XXI. Communications Systems Research

TELECOMMUNICATIONS DIVISION

A. Block Coding and Synchronization Study: Subcarrier Tracking Methods and Communication System Design, *W. C. Lindsey*

1. Introduction

Various communication systems, e.g., binary PSK, transmit information in the form $s(t) = (2)^{1/2} Am(t) \sin(\omega_0 t + \theta)$. In order that the received signal be demodulated coherently, it is necessary to determine or estimate the phase θ and frequency of the subcarrier $(2)^{1/2} \sin(\omega_0 t + \theta)$ with as little error as possible. If the signal $s(t)$ contains a residual component of sufficient strength at the subcarrier frequency, this component could be tracked with a narrow-band phase-locked loop and used to provide the desired reference signal. On the other hand, the power contained in the residual component represents power which does not convey any information other than the frequency and phase of the subcarrier. Thus, it represents power not available for the transmission of data, and, in practice, it is always of interest to investigate techniques which conserve and save power.

Several practical methods are available which rely upon the transmission of a reference signal. For example, the phase reference may be transmitted along with a PSK signal, and in order to maintain proper phase-synchronization, the phase-keyed signal and the reference signal must be close to each other in frequency and in time such that any channel fluctuations along the propagation path affect both signals the same way. For completeness, several methods of great practical importance are discussed below.

First, we have differential phase-shift keying (DPSK). In a DPSK system the PSK signal serves as the data signal and the reference signal. The phase of the signal received during one signaling interval serves as a reference for the next keying interval. The Kineplex (Ref. 1) is an example of a system which has been mechanized.

Second, we have the so-called adjacent tone reference PSK system (AT-PSK). The reference signal for this system is transmitted at an adjacent frequency simultaneously with the keyed signal. At the receiver, the phase of the reference is adjusted to compensate for the frequency difference between the reference signal and the phase-keyed signal. A practical system which employs this principle is illustrated by the DEFT system (Ref. 2)

Third is a system referred to as the quadrature reference PSK system (Q-PSK). In this system, the phase of one quadrature component is modulated with the data stream while the phase of the in-phase component remains unkeyed. The Kathryn system is an example of this technique (Ref. 3).

Finally, the so-called decision-directed measurement PSK (DDM-PSK) technique is employed. This system reconstructs a reference signal by estimating the modulation itself and using this estimate to eliminate it from the received signal. The decision directed system is, in essence, a generalization of the DPSK system, which uses the previous signaling interval. A computer simulation of this type of system has been carried out by Proakis, Drouilhet, and Price (Ref. 4). More recently Bussgang and Leiter (Ref. 5) derived the performance of a communication system in which a reference signal is transmitted at a frequency adjacent to the phase-keyed tone. Also, Bussgang and Leiter report results pertinent to the problem of the joint occurrence of two character errors on a multiple phase-keyed signal.

A number of methods have been proposed for generating a reference subcarrier from the received signal even when the residual subcarrier component is not available. This report analyzes and compares two methods of great practical interest in deep-space work. The results of the analysis are used to establish the performance of phase-coherent communication systems which utilize such subcarrier tracking methods. The first, the squaring-loop method, has been analyzed in a number of papers, Refs. 6–9. The second method, originally proposed by Costas, is the Costas-loop (Ref. 10). This article establishes the performance of these two subcarrier tracking methods using the Fokker-Planck apparatus as opposed to using linear tracking theory. The results are then used in predicting the performance of uncoded and block-coded communication systems. The theory developed is useful in the design and testing of subcarrier tracking loops and data detectors.

2. The Squaring-Loop Method

Of main concern here will be that of establishing a coherent subcarrier reference for demodulation of 180-deg PSK modulation. The mechanization of a typical squaring loop is illustrated in Fig. 1. The received signal $y(t)$ is bandpass filtered, squared to remove the modulation $m(t)$, and the resultant double frequency term is tracked by means of a conventional phase-locked loop (PLL) whose noise bandwidth is w_L cycles. When the output frequency of the *PLL* is divided by two, a coherent reference signal is available for demodulation purposes.

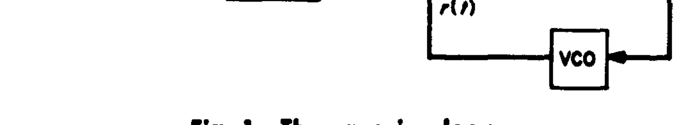

Fig. 1. The squaring loop

In deciding upon a method of determining the performance of the squaring loop, a significant parameter is the bandwidth of the bandpass filter whose transfer function is denoted by $H_i(p)$, and p is the Heaviside operator. In fact, if the input is contaminated by white noise of spectral density N_0 w cycle single-sided, and if the bandwidth of the filter is so large that the correlation time τ_i of its output noise is much smaller than the time constant $1/w_L$ of the phase-locked loop,[1] the squaring-loop may be analyzed by using the mathematical techniques available from the theory of Markov processes—in particular, the Fokker-Planck equation, Ref. 11.

As the bandwidth B_i of the bandpass filter is narrowed, the correlation time τ_i of the output noise increases and may become equal or even greater than the time constant $1/w_L$ of the phase-locked loop. The cases when $\tau_i \geqq 1/w_L$ are no less important in practice than the other extreme, when $\tau_i < 1/w_L$. However, the latter case is considered here for a constant frequency signal, and we neglect any spurious noise which may be generated due to imperfect system oscillators. Such fluctuations may be included with no great mathematical difficulty.

Let the observed data $y(t)$ be denoted by

$$y(t) = (2)^{1/2} A m(t) \sin(\omega_0 t + \theta) + n(t) \qquad (1)$$

where $m(t)$ is the signal envelope, i.e., the modulation, and let

$$n(t) = n_1(t)\cos(\omega_0 t + \theta) + n_2(t)\sin(\omega_0 t + \theta) \qquad (2)$$

be a realization of narrowband noise process, where $n_1(t)$ and $n_2(t)$ are sample functions of joint stationary Gaussian processes. We assume that the correlation time τ_n of the noise is small in comparison with the time constant of the PLL, i.e., $\tau_n << 1/w_L$.

Assuming a *perfect* square-law characteristic, the output process $z(t)$ is related to the input process $y(t)$ through

$$z(t) = [y(t) H_i(p)]^2 \qquad (3)$$

[1] Correlation time of the random process $\{x(t)\}$ is defined by the relation $\tau = \int_0^\infty |R_x(\tau)| d\tau$, where $R_x(\tau)$ is the normalized correlation function of the process. The parameter τ gives some idea of the size of the time interval over which correlation extends between values of the process $x(t)$.

where $H_i(p)$ is the transfer function of the bandpass filter and $p = d/dt$ is the Heaviside operator.

On substituting for $y(t)$ into Eq. (3) and taking only the terms around $2\omega_0$ yields, in operator form,

$$z(t) = H_i^2(p)\left\{\left[-A^2m^2(t) + \frac{n_1^2(t)}{2} - \frac{n_2^2(t)}{2} - (2)^{1/2}Am(t)\,n_2(t)\right]\cos(2\omega_0 t + 2\theta) + [(2)^{1/2}Am(t)\,n_1(t) + n_1(t)\,n_2(t)]\sin(2\omega_0 t + 2\theta)\right\} \quad (4)$$

The output of the multiplier is $\epsilon(t) = K_m z(t)\,r(t)$, where K_m is the multiplier constant. A convenient representation for $r(t)$ is denoted by

$$r(t) = (2)^{1/2}\sin[2\omega_0 t + 2\hat{\theta}] \quad (5)$$

If one takes only those terms in the base band frequency region, the product $r(t)\,z(t)$ becomes

$$z(t)\,r(t) = \frac{K_m H_i^2(p)}{2}\left\{\left[A^2m^2(t) - \frac{n_1^2(t) - n_2^2(t)}{2} + (2)^{1/2}Am(t)\,n_2(t)\right]\sin[2(\theta - \hat{\theta})] + [(2)^{1/2}Am(t)\,n_1(t) + n_1(t)\,n_2(t)]\cos[2(\theta - \hat{\theta})]\right\} \quad (6)$$

The phase $\hat{\theta}(t)$ of the voltage control oscillator (VCO) output is related to its input through

$$\hat{\theta}(t) = \frac{K_{VCO}}{p}\,z(t)\,r(t)\,F(p) \quad (7)$$

where K_{VCO} is the VCO gain constant in rad/sec-v. Neglecting any doppler present (this will be small in practice) on $\theta(t)$ we have from Eqs. (6) and (7) the following stochastic differential equation of operation of a squaring loop, viz.,

$$p\phi + \frac{K_m K_{VCO} A^2 m^2(t)\,H_i^2(p)\,F(p)\sin 2\phi}{2} = u(t, \phi) \quad (8)$$

where $\phi = \theta - \hat{\theta}$ and

$$u(t, \phi) = \frac{K_{VCO} K_m F(p)\,H_i^2(p)}{2}\left\{\left[\frac{n_1^2(t)}{2} - \frac{n_2^2(t)}{2} - (2)^{1/2}Am(t)\,n_2(t)\right]\sin 2\phi - [(2)^{1/2}Am(t)\,n_1(t) + n_1(t)\,n_2(t)]\cos 2\phi\right\} \quad (9)$$

If we let $\Phi = 2\phi$, $K = K_{VCO}K_m$, assume that over the bandwidth of significant interest that the filter $H_i(p) = 1$, and consider a first-order PLL, i.e., $F(p) = 1$, we have

$$\dot{\Phi} + KA^2m^2(t)\sin\Phi = K\left\{\left[\frac{n_1^2(t)}{2} - \frac{n_2^2(t)}{2} - (2)^{1/2}Am(t)\,n_2(t)\right]\sin\Phi - [(2)^{1/2}Am(t)\,n_1(t) + n_1(t)\,n_2(t)]\cos\Phi\right\} \quad (10)$$

We may now determine the probability density of Φ, using the Fokker-Planck method. The equation of operation is of the form $\dot{\Phi} = F[\Phi, u(\Phi, t)]$ for which the corresponding Fokker-Planck equation (Ref. 11) is, in the stationary case,

$$\frac{1}{2}\frac{\partial^2}{\partial\Phi^2}[K_2(\Phi)\,p(\Phi)] - \frac{\partial}{\partial\Phi}[K_1(\Phi)\,p(\Phi)] = 0 \qquad (11)$$

where

$$K_1(\Phi) = \overline{F[\Phi, u(\Phi, t)]}$$

and

$$K_2(\Phi) = \int_{-\infty}^{\infty} \{\overline{F[\Phi, u(\Phi, t)]\,F[\Phi, t + \tau)]} - K_1^2(\Phi)\}\,d\tau$$

and the bar denotes statistical averaging over the ensemble. If we make the assumptions that Φ is a slowly varying process, $m(t) = \pm 1$, we find from Eqs. (10) and (11) that

$$K_1(\Phi) = KA^2 \sin\Phi$$

and

$$K_2(\Phi) = K^2\sigma^2 \int_{-\infty}^{\infty} [\sigma^2 R_{n_1}^2(\tau) + 2A^2 R_{n_1}(\tau)]\,d\tau \qquad (12)$$

where σ^2 and $R_{n_1}(\tau)$ are, respectively, the variance and the envelope of the correlation function of the noise component in Eq. (2). They correspond to the variance and correlation function of the independent processes $n_1(t)$ and $n_2(t)$ in Eq. (2). Substitution of Eq. (12) into the partial differential equation given in Eq. (11) and using the boundary conditions

$$\int_{-\pi}^{\pi} p(\Phi)\,d\Phi = 1$$

$$p(\Phi + 2\pi) = p(\Phi) \qquad (13)$$

we have as a solution to Eq. (11)

$$p(\Phi) = \frac{\exp\left[\dfrac{AK}{K_2}\cos\Phi\right]}{2\pi I_0(AK/K_2)} \qquad |\Phi| < \pi \qquad (14)$$

where $I_0(x)$ is the modified Bessel function of zero order and of argument x. If we introduce the change of variable $\Phi = 2\phi$ and make use of the Jacobian of the transformation, we find that

$$p(\phi) = \frac{\exp[(AK/K_2)\cos 2\phi]}{\pi I_0(AK/K_2)}; \qquad |\phi| < \pi/2 \qquad (15)$$

As a first example of our results, assume that the normalized correlation function of the envelope of the input noise process possesses a Markov-type power spectrum with variance $\sigma^2 = N_0 B_i$, i.e.,

$$R_{n_1}(\tau) = \exp[-2B_i|\tau|] = R_{n_2}(\tau) \qquad (16)$$

where B_i is the one-sided bandwidth of the noise $n_1(t)$ or $n_2(t)$. Physically, Eq. (16) represents a noise source that has been generated by passing white noise through an RC filter which possesses a 3-db frequency of $B_i/2\pi$ Hz. Thus, $K_2(\phi)$ in Eq. (12) becomes

$$K_2 = 2K^2\sigma^2\left[\frac{\sigma^2}{4B_i} + \frac{A^2}{B_i}\right] \qquad (17)$$

and the solution in Eq. (15) is given by

$$p(\phi) = \frac{\exp[D\cos 2\phi]}{\pi I_0(D)}; \qquad |\phi| \leq \pi/2 \qquad (18)$$

where

$$D = \frac{x}{8}\left[\frac{1}{1 + \dfrac{32}{xy}}\right] \qquad (19)$$

and

$$x = \frac{2A^2}{N_0 w_L}; \qquad y = \frac{B_i}{w_L}$$

Here w_L is taken to be the bandwidth of the loop, as defined from the linear PLL theory, i.e.,

$$w_L = 2b_L = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} |H(s)|^2\,ds = A^2K/4 \qquad (20)$$

where $H(s)$ is the closed-loop transfer fnnction of the loop in linearized form. The square of the signal amplitude A is present because of the squaring operation.

As a second example, assume that

$$R_{n_1}(\tau) = R_{n_2}(\tau) = \frac{\sin \pi B_i \tau}{\pi B_i \tau} \tag{21}$$

then it is easy to show that

$$D = \frac{x}{4}\left[\frac{1}{1 + 1/xy}\right] \tag{22}$$

where

$$x = \frac{2A^2}{N_0 w_L}; \qquad y = \frac{w_L}{B_i}$$

Other presquaring filters may be easily evaluated. The two examples given represent results for the limiting cases of the class of Butterworth-type spectra.

If one assumes that ϕ is small, then the distribution of ϕ becomes Gaussian with variance

$$\sigma_\phi^2 = \tfrac{1}{4}\,\sigma_\Phi^2 = \left\{x\left[\frac{1}{1 + 1/xy}\right]\right\}^{-1} \tag{23}$$

This result agrees with that obtained using linear PLL theory (Refs. 1–4). The variance of the phase-error Φ, as determined from Eq. (14), is

$$\sigma_\Phi^2 = \frac{\pi^2}{3} + 4\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\frac{I_k(D)}{I_0(D)}$$

where $I_k(D)$ is the modified Bessel function of order k and argument D. For large D, σ_Φ^2 approaches $1/D$, as it should.

3. The Costas-Loop

In the Costas-loop shown in Fig. 2, the phase of the data subcarrier is extracted from the suppressed carrier signal $s(t)$ plus noise $n(t)$ by multiplying the input voltages of the two phase detectors (multipliers) with that produced from the output of the VCO and a 90-deg phase shift of that voltage, filtering the results and using this signal to control the phase and frequency of the loop's VCO output.

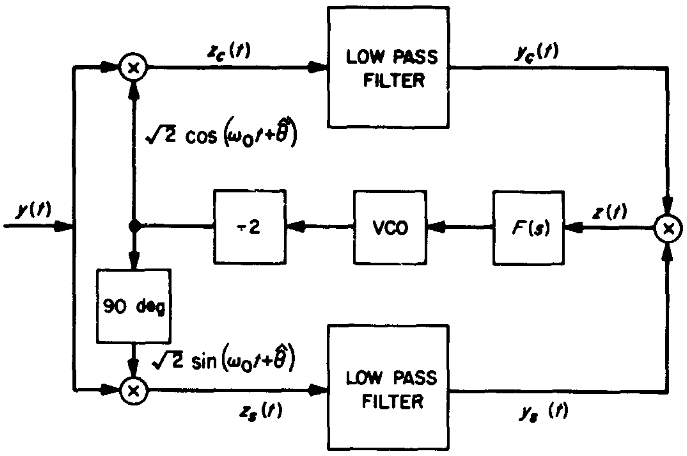

Fig. 2. The Costas-loop

If we denote the output of the upper loop multiplier by $z_c(t)$ and the output of the lower loop multiplier by $z_s(t)$ (see Fig. 2) then the output $z_c(t)$ is

$$z_c(t) = y(t)\cdot(2)^{1/2}\cos(\omega_0 t + \hat{\theta}) \tag{24}$$

while the output of the low-pass filter becomes

$$y_c(t) = \left[Am(t) + \frac{n_2(t)}{(2)^{1/2}}\right]\sin\phi + \frac{n_1(t)}{(2)^{1/2}}\cos\phi \tag{25}$$

when Eq. (1) is substituted into Eq. (24) and all double frequency terms are neglected. Similarly, the output $y_s(t)$ is given by

$$y_s(t) = \left[Am(t) + \frac{n_2(t)}{(2)^{1/2}}\right]\cos\phi - \frac{n_1(t)}{(2)^{1/2}}\sin\phi \tag{26}$$

The control voltage $z(t) = y_c(t)\,y_s(t)$ becomes

$$\begin{aligned} z(t) &= \tfrac{1}{2}\left(Am(t) + \frac{n_2(t)}{(2)^{1/2}}\right)^2\sin 2\phi \\ &+ \frac{n_1(t)}{(2)^{1/2}}\left(Am(t) + \frac{n_2(t)}{(2)^{1/2}}\right)\left(\frac{1+\cos 2\phi}{2}\right) \\ &- \frac{n_1(t)}{(2)^{1/2}}\left(Am(t) + \frac{n_2(t)}{(2)^{1/2}}\right)\left(\frac{1-\cos 2\phi}{2}\right) \\ &- \frac{n_1^2(t)}{4}\sin 2\phi \end{aligned} \tag{27}$$

Now

$$\hat{\theta} = K_{VCO} K_m F(p) \cdot z(t) \tag{28}$$

and if we omit all dc terms, the stochastic differential equation which governs the behavior of the Costas-loop in the presence of noise reduces to

$$\Phi + KA^2 F(p) m^2(t) \sin \Phi = KF(p) \left\{ \left[\frac{n_1^2(t)}{2} - \frac{n_2^2(t)}{2} - (2)^{1/2} Am(t) n_2(t) \right] \sin \Phi - \left[(2)^{1/2} An_1(t) m(t) + n_1(t) n_2(t) \right] \cos \Phi \right\} \tag{29}$$

If, in the previous case, we ignore the effects of the filter $H_i(p)$, then the stochastic differential equation obtained for the squaring-loop method and the stochastic differential equation for the Costas-loop are identical. Thus, the solution for $p(\Phi)$ is identical, and the noise behavior of the two circuits is the same. This, of course, assumes that the low-pass filter transfer functions can be obtained by simply translating, by f_0 Hz, the bandpass filter function of the squaring-loop.

From this it may be concluded that the two approaches to subcarrier tracking yield equivalent results when the filters in the Costas-loop are the low-pass equivalents of the bandpass filter in the squaring-loop. The choice of which loop to use cannot be determined on theoretical grounds, and consequently, must be determined from an engineering hardware point of view, i.e., the relative ease with which the corresponding filters can be constructed. Both methods of subcarrier tracking exhibit the usual 180-deg phase ambiguity inherent in all systems that attempt to recover the subcarrier phase from a modulated signal, i.e., changing the sign of the received signal leaves the sign of the recovered subcarrier unaltered.

An obvious question coming to mind is to ask for the presquaring filter which maximizes the signal-to-noise ratio (SNR) at the output of the phase-locked loop. This problem has been solved, and the optimum filter, for the case where the modulating spectrum is narrow with respect to the carrier frequency, has been shown (SPS 37-37, Vol. IV, p. 290) to be given by

$$H_i(p) = k \left(\frac{S_s(p)}{S_s(p) + N_0/2} \right)^{1/2} \tag{30}$$

where k is an arbitrary positive constant and $S_s(p)$ is the power spectrum of modulated signal $s(t)$. For large signal-to-noise conditions the optimum presquaring filter given by Eq. (30) becomes $H_i(s) \sim k$ while for small signal-to-noise conditions the optimum filter becomes

$$H_i(p) \sim \left(\frac{2S_s(p)}{N_0} \right)^{1/2} \tag{31}$$

This says that for small signal-to-noise conditions the optimum filter is matched to the signaling spectrum. Arbitrarily setting $k = 1$ says that the optimum filter for large SNR is an ideal bandpass filter for which the performance has been accessed. On the other hand, for small SNR the performance of the two loops may be accessed once the spectrum of the modulated signal $s(t)$ is defined. It is our conjecture that the improvement over an ideal bandpass filter is negligibly small in the SNR region where such synchronization techniques are useful in practice. In the next section we show that squaring-loops or Costas-loops are most useful in data detection systems where the ratio of data rate $\mathcal{R}$ to the tracking loop bandwidth w_L is large, i.e., high data rate systems.

Various other approaches to the problem of estimating the subcarrier phase when no residual component is present at the subcarrier frequency are available, and in some cases have been analyzed. Layland (Ref. 9) and Proakis (Ref. 4) analyze methods which essentially estimate the modulation itself. This estimate is used in an attempt to eliminate the modulation from the subcarrier. This, therefore, provides an unmodulated sinusoid which can be tracked by a PLL.

4. Performance of Correlation Receivers

Consider the situation where $\{m(t)\}$ represents the set of signals $\{x_k(t), k = 1, \cdots, N\}$. For the present we assume that each signal in the set occurs with equal probability, contains equal energies, exists for a time duration of T seconds and is orthogonal, i.e.,

$$\int_0^T x_k(t)\,x_j(t) = \delta_{jk} \tag{32}$$

where $\delta_{jk} = 1$ for $j = k$ and $\delta_{jk} = 0$ for $j \neq k$. In the presence of white Gaussian noise the optimum receiver, i.e., the one which minimizes the error probability, computes

$$C_k = \int_0^T y(t)\,x_k(t)\,dt \tag{33}$$

for all $k = 1, \cdots, N$ and makes its decision in favor of that signal which yields the largest C_k.

Of particular interest here is the case where the set of signals $\{x_k(t)\}$ are code words taken from an orthogonal code dictionary containing $N = 2^n$ code words, i.e., the signals are sequences of + and − 1's. In this case the time duration T becomes the product of the number of bits per code word times the time duration per bit, i.e., $T = nT_b$. If one assumes that word sync and symbol sync[2] (i.e., the instants in time where one word begins and another ends and the instants in time where the modulation may change states) are known exactly and that either the squaring-loop method or Costas-loop is used to provide subcarrier sync, the conditional probability that the data detector will err may be shown to be given by (Refs. 13 and 14)

$$P_E(\phi) = 1 - P_c(\phi) = 1 - \int_{-\infty}^{\infty} \frac{\exp(-y^2/2)}{(2\pi)^{1/2}}\,dy \left[\int_{-\infty}^{y+(2Rn)^{1/2}\cos\phi} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{x^2}{2}\right) dx\right]^{2^n-1} \tag{34}$$

where $R = A^2T_b/N_0$. The average word error probability is obtained from Eq. (34) by averaging over the distribution of $p(\phi)$, i.e.,

$$P_E = \int_{-\pi/2}^{\pi/2} p(\phi)\,P_E(\phi)\,d\phi \tag{35}$$

Thus, from Eqs. (18), (34) and (35) we have[3]

$$P_E = 1 - \int_{-\pi/2}^{\pi/2} \frac{\exp(D\cos 2\phi)}{\pi I_0(D)}\,d\phi \int_{-\infty}^{\infty} \frac{\exp(-y^2/2)\,dy}{(2\pi)^{1/2}} \left[\int_{-\infty}^{y+(2nR)^{1/2}\cos\phi} \frac{\exp(-x^2/2)}{(2\pi)^{1/2}}\,dx\right]^{2^n-1} \tag{36}$$

where

$$D = \frac{\delta y R}{4}\left[\frac{1}{1+\dfrac{1}{\delta y R}}\right] \tag{37}$$

if an ideal bandpass filter precedes the squaring-loop or

$$D = \frac{\delta y R}{8}\left[\frac{1}{1+\dfrac{32}{\delta y R}}\right] \tag{38}$$

if an RC filter centered around ω_0 precedes the squaring-loop. The parameters R, δ, y and $\mathcal{R}$ are defined by

$$R = \frac{A^2T_b}{N_0}; \qquad \sigma = \frac{2\mathcal{R}}{w_L}; \qquad y = \frac{w_L}{B_i}; \qquad \mathcal{R} = \frac{1}{T_b} \tag{39}$$

We point out that the parameter $\mathcal{R}$ is the data rate of the system.

[2] This assumption is not too restrictive since jitter on the phase of the subcarrier is more deleterious on system performance than jitter about those instants in time with which the modulation may change states. This, of course, is a consequence of coherent detection.

[3] In some cases the bit-error probability is of interest. The ratio of the bit-error probability to the word error probability is $2^{n-1}/2^n - 1$ (Ref. 14).

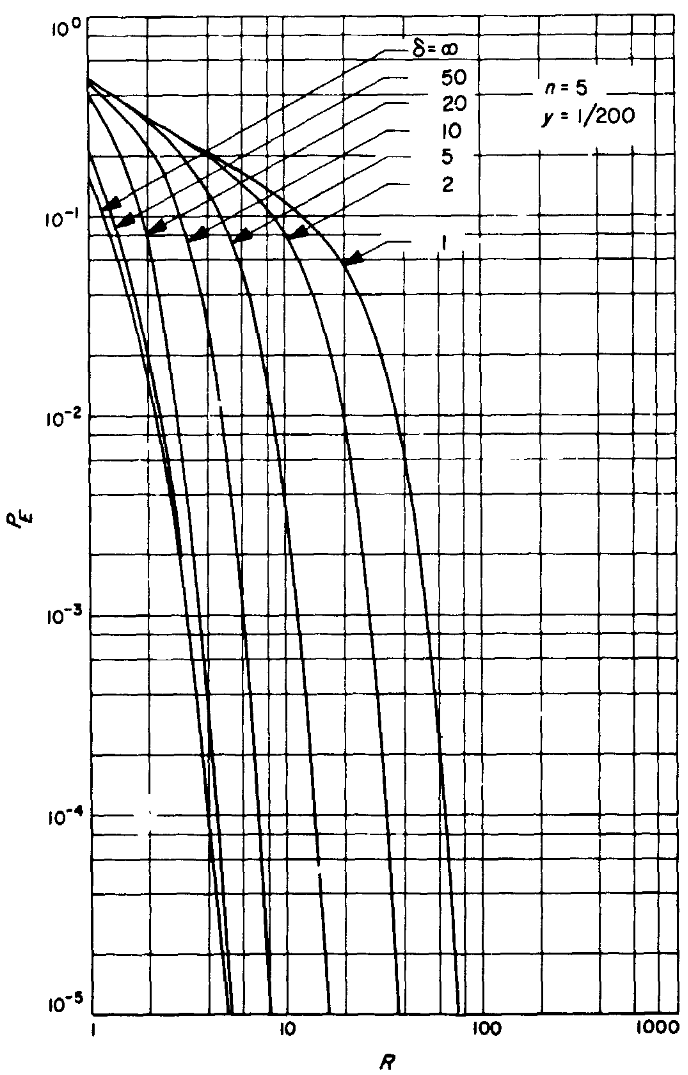

Fig. 3. Word error probability P_E vs R ($n = 5$)

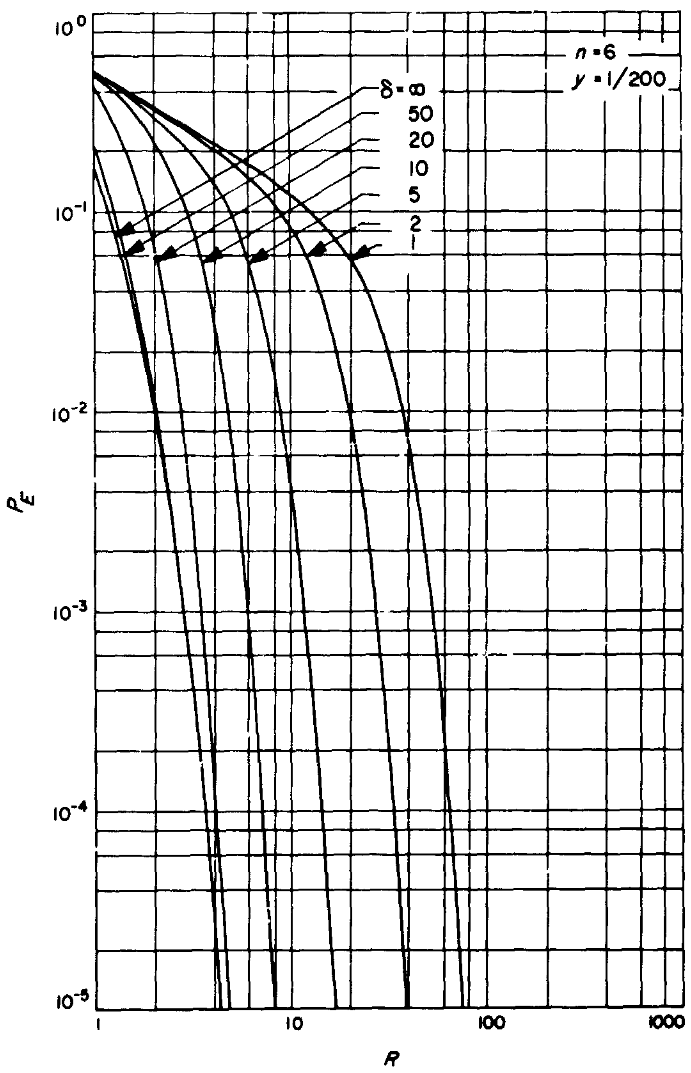

Fig. 4. Word error probability P_E vs R ($n = 6$)

Eq. (36) has been integrated numerically using an IBM 704 computer for $n = 5$–8. The results of the numerical integration are illustrated in Figs. 3–6 for the situation where an ideal bandpass filter precedes the PLL. The parameter y was set at 1/2000, since this is typical of what might be encountered in practice. It is clear from these figures that obtaining subcarrier sync by the method outlined here is most beneficial in systems where $\delta = 2\mathcal{R}/w_L >> 1$, i.e., high data rate systems.

One may proceed to develop the performance of such a system for biorthogonal codes. However, if one recalls that for $n \geqq 5$, the performance of systems which utilize orthogonal code dictionaries is approximately equal to systems which employ biorthogonal code dictionaries; then the results presented in Figs. 3–6 may be used in carrying out a particular design where biorthogonal codes are used. The performance of a block-coded system which uses biorthogonal codes is given by

$$P_E = 1 - \int_{-\pi/2}^{\pi/2} P_c(\phi) \frac{\exp[D\cos 2\phi]}{\pi I_0(D)} d\phi \tag{40}$$

where

$$P_c(\phi) = \int_{-(2nR)^{1/2}\cos\phi}^{\infty} \frac{\exp(-y^2/2)}{(2\pi)^{1/2}} \times \left[\int_{-y+(2nR)^{1/2}\cos\phi}^{y+(2nR)^{1/2}\cos\phi} \frac{\exp(-x^2/2)}{(2\pi)^{1/2}} dx\right]^{2^{n-1}-1} dy \tag{41}$$

For $n \geqq 5$, numerical integration of Eq. (38) on the IBM 7090 produces results, for all practical purposes, equivalent to those shown in Figs. 3–6.

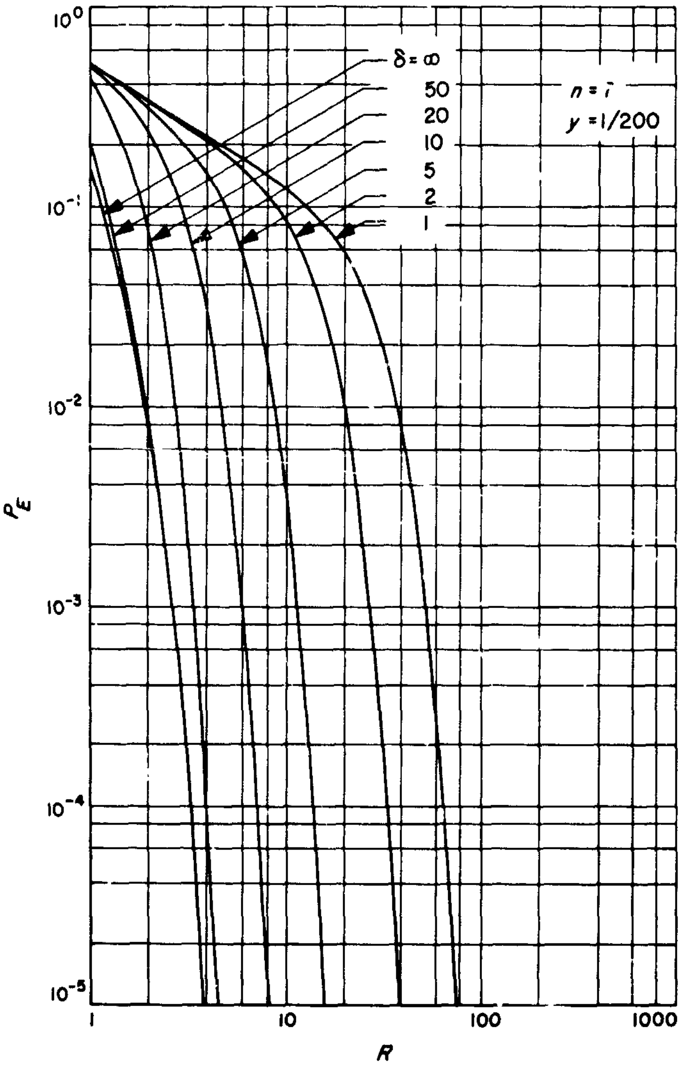

Fig. 5. Word error probability P_E vs R ($n = 7$)

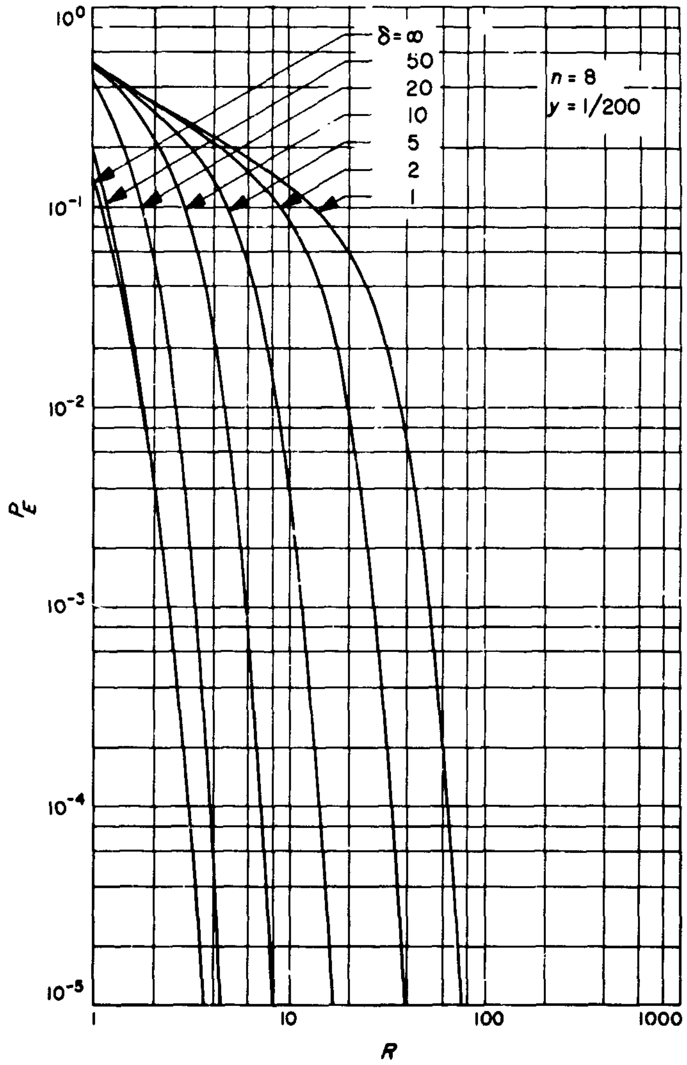

Fig. 6. Word error probability P_E vs R ($n = 8$)

Finally, it is of interest to understand how the value of $y = B_i/w_L$ affects the performance of a particular design. This trend is best illustrated, for various values of y, in Figs. 7–9 for uncoded telemetry systems, and in Figs. 4, 10 and 11 for block-coded telemetry systems. The results given in Figs. 7–9 were obtained by numerical integration of Eq. (40) with $n = 1$, while the results given in Figs. 4, 10 and 11 are, for all practical purposes, valid for biorthogonal codes, even though they were computed from Eq. (34). This is due to the fact, mentioned earlier, that for $n \geqq 5$ the performance of telemetry systems which employ orthogonal codes is approximately equivalent to that of telemetry systems which employ biorthogonal codes (Refs. 13 and 14). An obvious conclusion, which may be reached here, is that for a fixed δ and R system, performance improves as the ratio $y = B_i/w_L$ becomes larger. This result is comprehensible from a physical point of view.

5. Conclusion

A model probability distribution for the phase-error exhibited by the squaring-loop or Costas-loop has been derived using the Fokker-Planck equation. The parameters of this distribution are evaluated in terms of the covariance function of the input noise and, in particular, for two specific noise spectra. The model distribution is then used to assess the degradation in performance of a coded or uncoded telemetry system which tracks the phase of the subcarrier, using this method. If the phase of a suppressed carrier signal is derived from the modulated data subcarrier by means of a Costas-loop or a squaring-loop, the critical design parameter, which indicates the usefulness of such tracking loops in the demodulation process, is the ratio of the data-rate to the bandwidth of the loop. In the case of coded systems, this implies high-data rates for error rates less than 10^{-3}.

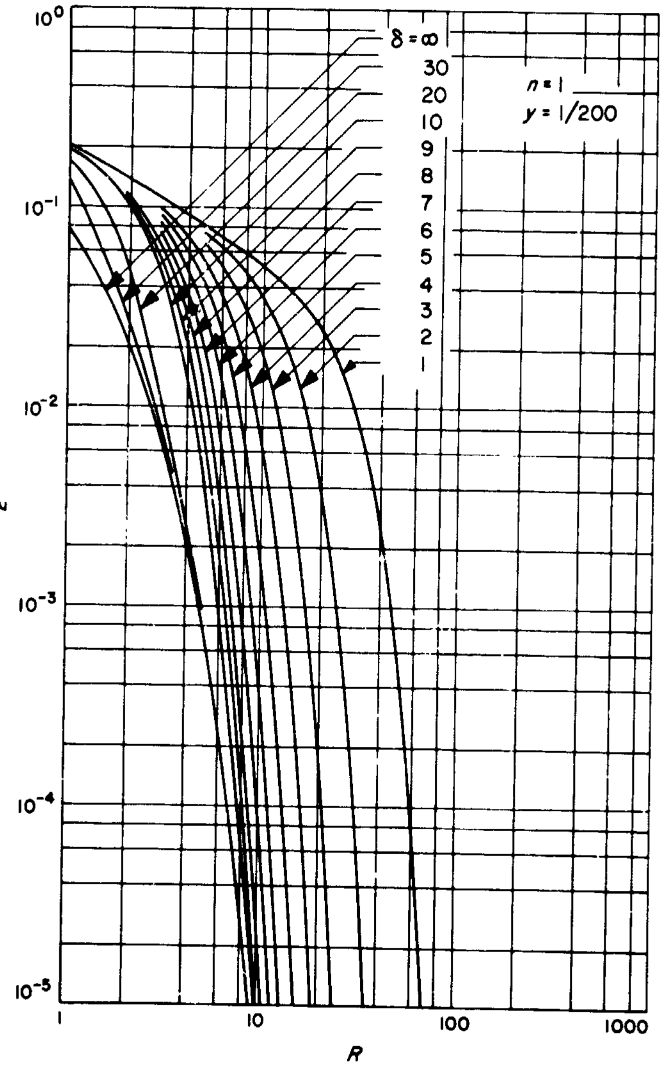

Fig. 7. Bit error probability P_E vs R ($n = 1, y = 1/200$)

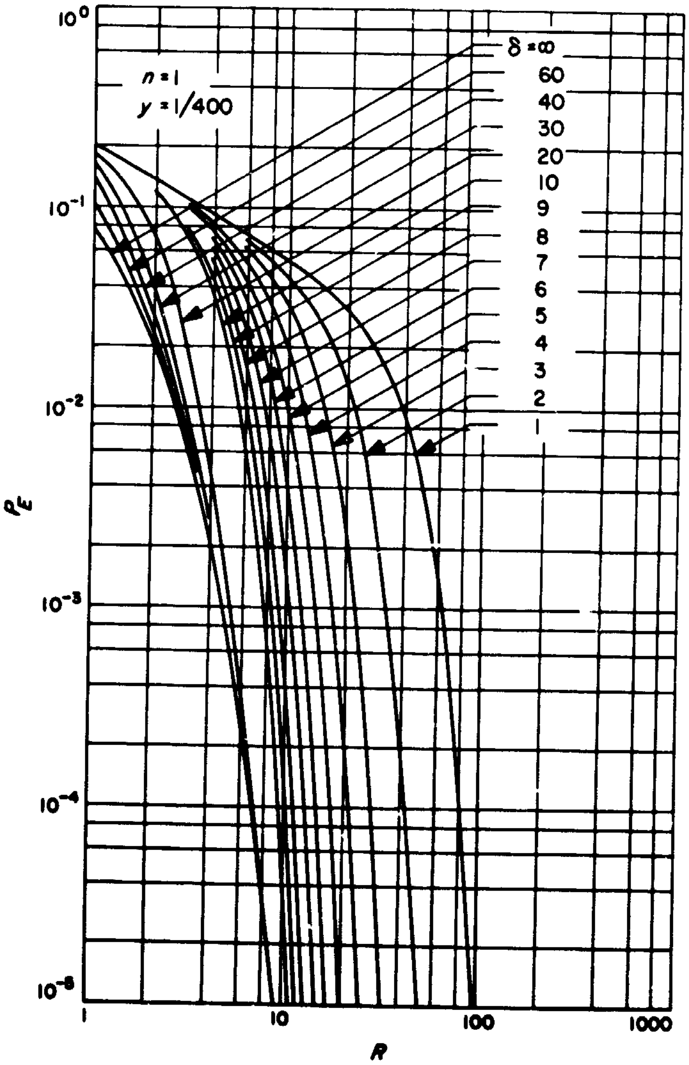

Fig. 8. Bit error probability P_E vs R ($n = 1, y = 1/400$)

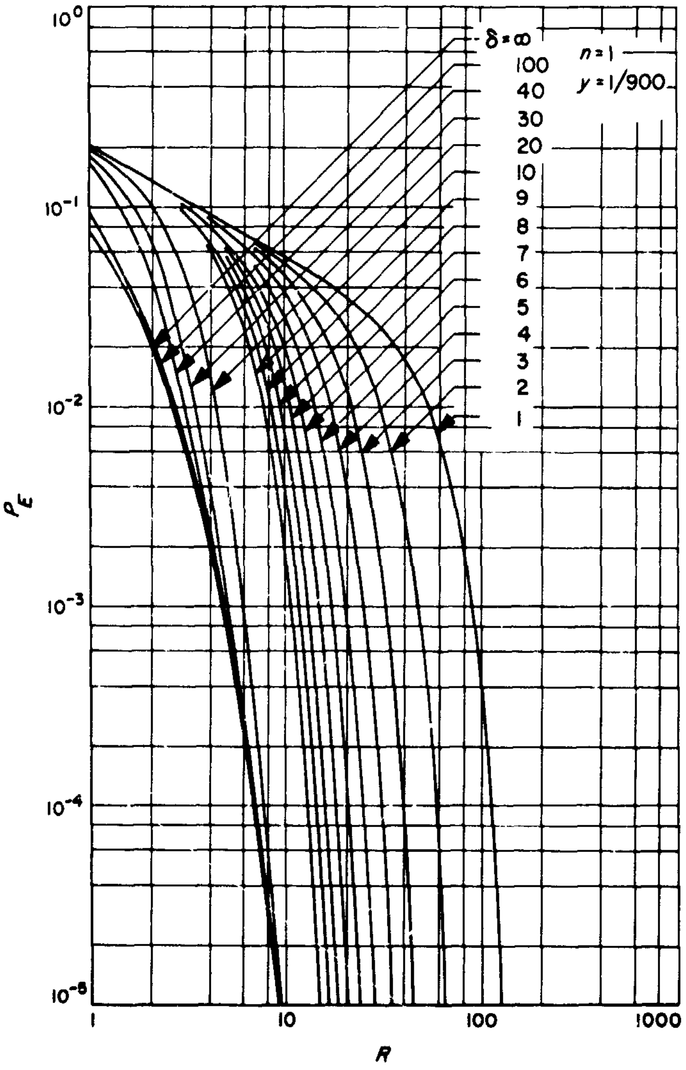

Fig. 9. Bit error probability P_E vs R ($n = 1, y = 1/900$)

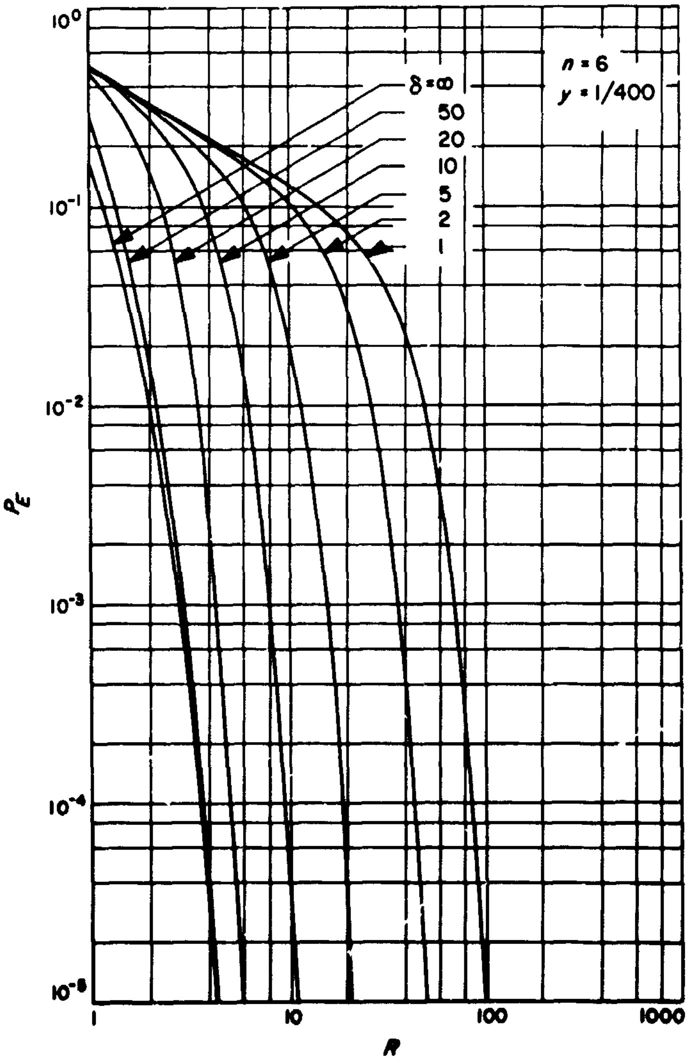

Fig. 10. Word error probability P_E vs R ($n = 6, y = 1/400$)

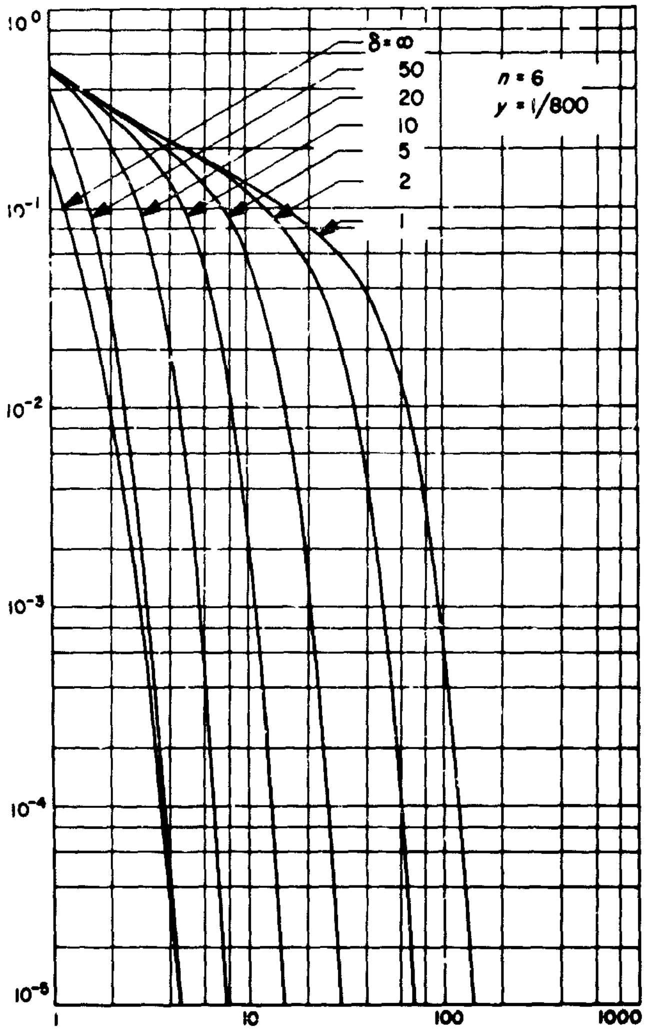

Fig. 11. Word error probability P_E vs R ($n = 6, y = 1/800$)

B. Block Coding and Synchronization Study: Power Allocation and Noisy Reference Losses in Phase Coherent Communication Systems, *W. C. Lindsey*

1. Introduction

The problem of power allocation in deep-space telemetry and command systems is becoming more important in the design and engineering of communication systems. The reason, of course, is that it is necessary to be able to predict accurately, prior to launch, the actual behavior and performance of the system at various times after launch. If this can be accomplished with precision, engineering tradeoffs may be recognized, and the cost of the mission may be minimized with respect to mission yield. This is becoming particularly evident in the operation of a complex communication network, such as the deep space network.

In this article we present a method whereby the total transmitter power may be optimally allocated in a single-channel, phase-coherent communication system of the type discussed in Ref. 14. The novelty of the method lies in the fact that it is simple and can be carried out without the aid of a general purpose digital computer. The method takes into consideration the radio-frequency (RF) carrier phase-jitter due to a noisy RF reference. The power is allocated on the basis of minimizing the probability that the data detector will err in making its decision. Other results are given which enable one to determine the losses due to noisy demodulation references in one-way and two-way systems. The symbols used in our calculations are defined in Table 1.

2. Basic System Model

In order to shorten the subsequent derivation, we draw heavily upon previous results and the notation given in Refs. 14–16. The basic form of a two-way coherent communication link is depicted in Fig. 12. Briefly,[4] the reference transmitter phase modulates the RF carrier, say

[4] Reference 14 gives a more detailed description of the overall system.

Table 1. Definition of symbols

Symbol	Definition
$n = 1$	One-way communications
$n = 2$	Two-way communications
P_n	Total average transmitted power
T_{bn}	Time duration per bit
$\mathcal{R}_n$	System data rate
N_{0n}	Noise spectral density (single-sided)
ϕ_n	System phase-error
$F_n(s)$	Carrier tracking loop filter
w_{Ln}	Bandwidth of carrier tracking loop
r_n	Second-order loop parameter ratio $(P)^{1/2} K \tau_2^2/\tau_1$
$H_n(s)$	Closed loop transfer function of the carrier tracking loop assuming linear PLL theory
δ_n	Ratio of system data rate to carrier tracking loop bandwidth
R_n	Total signal power-to-noise spectral density ratio times the duration per bit
G	Static phase gain
β	Ratio of carrier tracking loop bandwidth in the vehicle system to that in the reference system
$P_E(n)$	System error probability
$P_{E0}(n)$	Minimum system error probability
R_{d1}	SNR in data (one-way link)
R_{d2}	SNR in data (two-way link)
α	SNR in carrier tracking loop

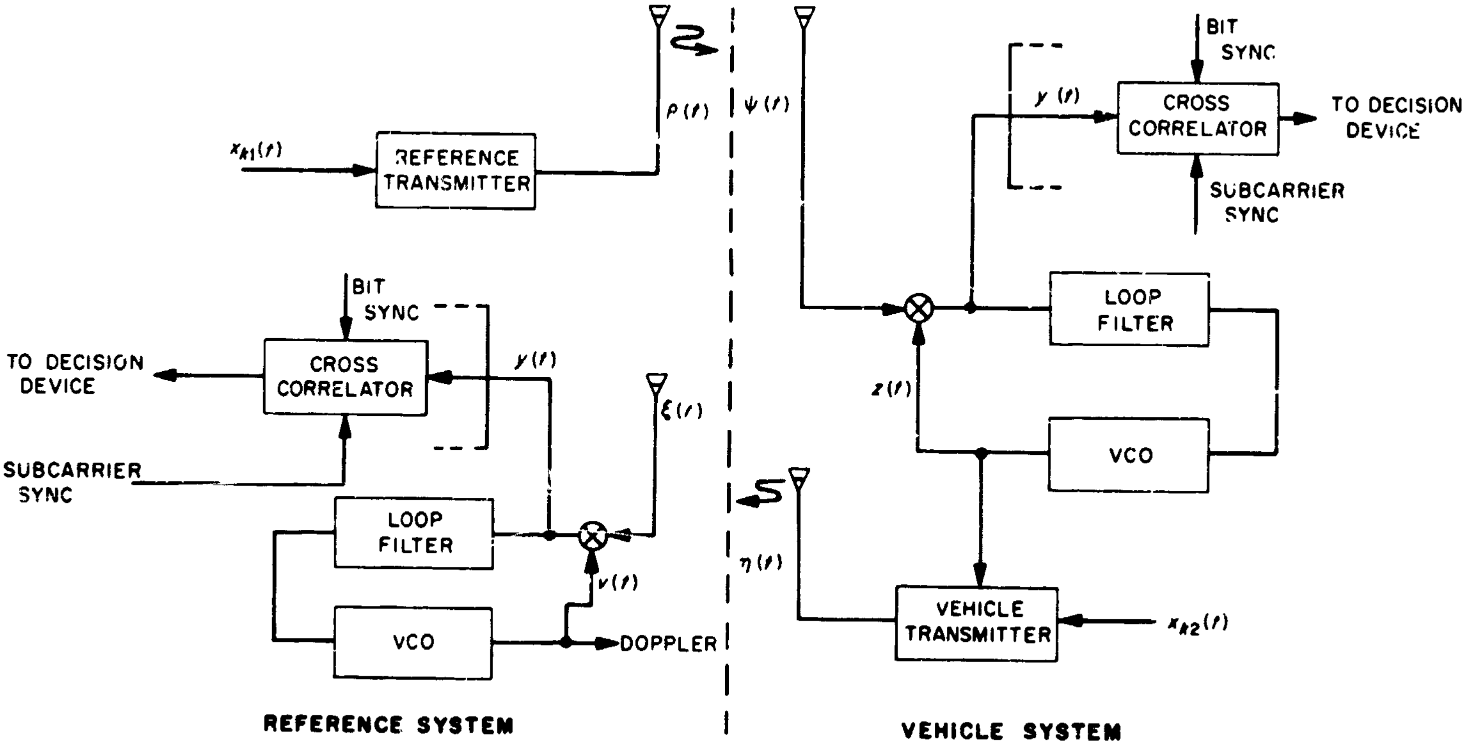

Fig. 12. Two-way communication link

$c(t)$, with one of two equally probable, equienergy signals $x_{k1}(t)$, $k = 1, 2$ of time duration T_{b1} sec. This signal is conveniently denoted by

$$\rho(t) = (2P_1)^{1/2} \sin[\omega t + (\cos^{-1} m_1) x_{k1}(t)] \tag{1}$$

where P_1 is the total radiated power, and m_1 is the modulation factor associated with the reference system. The channel introduces an arbitrary (but unknown) phase shift in the transmitted waveform and further disturbs $\rho(t)$ with additive white Gaussian noise $n_1(t)$ of single-sided spectral density of N_{01} w/Hz. Thus, we observe in the vehicle the following signal:

$$\psi(t) = (2P_1)^{1/2} \sin[\omega_1 t + (\cos^{-1} m_1) x_{k1}(t) + \theta_1] + n_1(t) \tag{2}$$

The vehicle tracks the carrier component in $\psi(t)$ by means of a narrow band phase-locked loop (PLL). The output $z(t)$ of the voltage control oscillator (VCO) of this tracking filter is used as a coherent reference in demodulating $\psi(t)$. The vehicle reference waveform $z(t)$ is conveniently taken to be

$$z(t) = (2)^{1/2} \cos(\omega_1 t + \hat{\theta}_1) \tag{3}$$

where $\hat{\theta}_1$ is the PLL estimate of θ_1 in the presence of noise. After neglecting the double frequency components and assuming that the data biphase modulates a square wave subcarrier, it may be shown (Ref. 14) that multiplication of $\psi(t)$ with $z(t)$ produces

$$y(t) = [(1 - m_1^2) P_1]^{1/2} x_{k1}(t) \cos\phi_1 + n_1'(t) \tag{4}$$

where $n_1'(t)$ is white Gaussian noise of single-sided spectral density of N_{01} w/Hz, and $\phi_1 = \theta_1 - \hat{\theta}_1$ is the vehicle subsystem phase error. We assume that this phase error is constant for at least T_{b1} sec. Also, we point out (Ref. 14) that m_1 represents the square root of the ratio of the power remaining in the carrier component to the total power radiated, i.e., $m_1 = (P_{c1}/P_1)^{1/2}$.

The decision in the vehicle is made in favor of that signal which gives rise to the largest cross correlation, i.e., the vehicle demodulator computes

$$q = \int_0^{T_{b1}} y(t) [x_{11}(t) - x_{21}(t)] \, dt \tag{5}$$

and compares the result with zero. If $q > 0$, x_{11} is announced, and if $q < 0$, x_{21} is announced.

In the reverse direction, i.e., transmission of the data back to the reference subsystem, the output of the vehicle's VCO is used as a carrier for transmission of one of two equienergy, equiprobable waveforms $x_{k2}(t)$, $k = 1,2$ of time duration T_{b2} sec. In this case, the output of the vehicle is conveniently represented by the following waveform

$$\eta(t) = (2P_2)^{1/2} \sin[\omega_2 t + (\cos^{-1} m_2) x_{k2}(t) + \hat{\theta}_1] \qquad (6)$$

Here m_2 is the modulation factor which represents the square root of the ratio of the power in the carrier to the total power radiated, i.e., $m_2 = (P_{c2}/P_2)^{1/2}$, (Ref. 14).

It is clear that using the vehicle VCO output as a carrier reference introduces into the down-link an additional component of noise; however, incorporating this measurement into the system allows one to perform, with extreme accuracy, a two-way doppler measurement. Thus, we postulate a mathematical model of the system so as to include this up-link jitter component, hence, the two-way doppler measurement. However, as we shall see, adjustment of certain parameters will immediately alleviate this up-link RF jitter. The down-link channel (assumed to be statistically independent from the up-link channel) further perturbs $\eta(t)$ by inserting an unknown phase shift θ_2 and additive white Gaussian noise $n_2(t)$ of single-sided spectral density of N_{02} w/Hz. Thus, the reference receiver observes

$$\xi(t) = (2P_2)^{1/2} \sin[\omega_2 t + (\cos^{-1} m_2) x_{k2}(t) + \hat{\theta}_1 + \theta_2] + n_2(t) \qquad (7)$$

The ground receiver tracks the carrier component in $\xi(t)$ for the purpose of measuring the doppler and demodulating the data. We denote the output of the reference VCO by

$$\nu(t) = (2)^{1/2} \cos(\omega_2 t + \hat{\theta}_2) \qquad (8)$$

where $\hat{\theta}_2$ is the estimate of phase of the observed carrier component. Multiplying $\xi(t)$ by $\nu(t)$ and neglecting the double-frequency components, one may show (Ref. 14) that

$$y(t) = ((1 - m_2^2) P_2)^{1/2} x_{k2}(t) \cos\phi_2 + n_2'(t) \qquad (9)$$

where $\phi_2 = \theta_2 + \hat{\theta}_1 - \hat{\theta}_2$ is the reference system phase error and is assumed constant for the duration T_{b2} of the signals $x_{k2}(t)$. Again $n_2'(t)$ is easily shown to be white with a single-sided spectral density of N_{02} w/Hz.

To recapitulate, we see that the design engineer has at his disposal several communication parameters. For the up-link we have the total power radiated P_1, the single-sided noise spectral density N_{01}, up-link data rate $\mathcal{R}_1 = T_{b1}^{-1}$, vehicle carrier tracking loop bandwidth w_{L1}, and modulation index $m_1 = (P_{c1}/P_1)^{1/2}$. The corresponding down-link parameters are P_2, N_{02}, $\mathcal{R}_2 = T_{b2}^{-1}$, w_{L2}, and $m_2 = (P_{c2}/P_2)^{1/2}$. In the subsections that follow, we relate these parameters together and determine that value of m_n $(n = 1, 2)$ which minimizes the probability of error $P_E(n)$ $(n = 1, 2)$, say $P_{E_0}(n)$, for a fixed data rate-to-carrier tracking loop bandwidth ratio, say $\mathcal{R}_n/w_{Ln}$ $(n = 1, 2)$. Also the losses, in signal-to-noise ratios (SNR), due to noisy demodulation references are determined.

3. System Phase-Error Distribution

The probability distribution for the subsystem phase error is of great importance in specifying the performance of the two-way link. In fact, the distribution of this phase-error has been previously characterized (Ref. 14), and its probability density function is conveniently represented by

$$P(\phi_2) = \frac{I_0(|\alpha_1 + \alpha_2 \exp(j\phi_2)|)}{2\pi I_0(\alpha_1) I_0(\alpha_2)}, \qquad |\phi_2| < \pi \qquad (10)$$

where $I_0(x)$ is the imaginary Bessel of zero order and argument x, and α_1 and α_2 are related to the up- and down-link parameters. The validity of using this distribution as a model for the phase-error distribution has been given in Refs. 14 and 15.

In passing we point out that the loop filter, which is of greatest interest in practice, is of the proportional-plus-integral control type,[5] i.e.,

$$F_n(s) = \frac{1 + \tau_{2n}s}{1 + \tau_{1n}s}; \qquad n = 1,2 \qquad (11)$$

If one relates the basic parameter of the carrier tracking loops, i.e., the loop bandwidth w_{Ln}, $n = 1, 2$, to the time

[5] The subscript $n = 1$ refers to parameters, filters, etc. in the vehicle system or one-way operation; while $n = 2$ refers to parameters, filters, etc. in the reference system.

constants in $F_n(s)$, we have

$$w_{Ln} = 2B_{Ln} = \frac{r_n + 1}{2\tau_{2n}}; \qquad r_n = \frac{(P_n)^{1/2} K_n \tau_{2n}^2}{\tau_{1n}} \tag{12}$$

where K_n is the equivalent simple-loop gain which depends upon the VCO constant and the multiplier constant, Ref. 16, p. 30. The loop bandwidth B_{Ln} is defined as

$$w_{Ln} = 2B_{Ln} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} |H_n(s)|^2\, ds; \qquad n = 1, 2 \tag{13}$$

where

$$H_n(s) = \frac{1 + \left(\dfrac{r_n + 1}{4B_{Ln}}\right)s}{1 + \left(\dfrac{r_n + 1}{4B_{Ln}}\right)s + \dfrac{1}{r_n}\left(\dfrac{r_n + 1}{4B_{Ln}}\right)^2 s^2}; \qquad n = 1, 2 \tag{14}$$

and $H_n(s)$, $n = 1, 2$ is the closed loop transfer function of the carrier tracking loops in linearized form. The transfer function of the ground receiver's carrier tracking loop is given by Eq. (14) with $n = 2$, and that of the spacecraft's carrier tracking loop is given by Eq. (14) with $n = 1$.

The parameters α_1 and α_2, which serve to characterize $p(\phi_2)$, aıe given by

$$\alpha_1 = \frac{2m_1^2 P_1}{N_{01} w_{L1}} \cdot \frac{1}{K(r_1, r_2, \beta)} \qquad \alpha_2 = \frac{2m_2^2 P_2}{N_{02} w_{L2}} = m_2^2 \delta_2 R_2 \tag{15}$$

where

$$\delta_2 = \frac{\mathcal{R}_2}{w_{L2}} = \frac{1}{T_{b2} w_{L2}} \qquad R_2 = \frac{2P_2 T_{b2}}{N_{02}} \tag{16}$$

defined by

$$K(r_1, r_2, \beta) = \frac{r_1 \beta G^2}{(r_1 + 1) r_2} \times \left[\frac{r_2 + r_1 r_2 \beta (1 + \beta) + \beta^2 (1 + \beta) + r_1 \beta^3}{r_2/r_1 + r_2 \beta + \beta^2 (r_2 + r_1 - 2) + r_1 \beta^3 + r_1 \beta^4 / r_2}\right] \tag{17}$$

with

$$\beta = \frac{B_{L1}(r_2 + 1)}{B_{L2}(r_1 + 1)}$$

and G is the static phase gain of the spacecraft receiver, which is determined by the ratio of the input frequency to the output frequency at the spacecraft. In practice, the values of β and G are chosen such that $K(r_1, r_2, \beta)$ is approximately unity; hence, α_1 is approximately the SNR existing in the carrier tracking loop bandwidth.

4. Losses Due to Noisy Demodulation References in One-Way and Two-Way Systems

Before proceeding with the problem of power allocation, we determine the RF losses due to a noisy carrier reference in one-way and two-way telemetry links. For two-way links, the significant contributing factors in this loss are the noise in the ground receiver's demodulation reference $\hat{\theta}_2$ and the phase modulation existing on the vehicles carrier produced by the up-link additive noise.

a. Losses in one-way links. One-way telemetry reception, i.e., reception when the vehicle is operating with an auxiliary oscillator as a frequency reference, has been presented and discussed in Refs. 17 and 18. For the sake of completeness, we include a graph (Fig. 13) which enables the design engineer to evaluate such losses when the reference phase-error is constant over a bit period. These results are obtained by numerically integrating the expression, which specifies the bit-error probability $P_E(1)$ as a function of the SNR in the data, say R_{d1}, and the SNR in the reference system's carrier tracking loop. This bit-error probability has been shown, Ref. 14, to be given by

$$P_E(1) = \lim_{\substack{\alpha_2 \to \infty \\ w_{L2} \to 0}} \int_{-\pi}^{\pi} p(\phi_2)\, \mathrm{Erfc}\left[(2R_{d1})^{1/2} \cos\phi_2\right] d\phi_2 \tag{18}$$

where

$$\mathrm{Erfc}(x) = \frac{1}{(2\pi)^{1/2}} \int_r^{\infty} \exp(-y^2/2)\, dy \tag{19}$$

Eq. (18) has been plotted in Fig. 14 for various values of the parameters $R_{d1} = (1 - m_1^2) R_1$ and α. Here $\alpha = 2P_{c1}/N_{01}w_{L1} = 2m_1^2 P_1 / N_{01} w_{L1}$ denotes the SNR in the carrier tracking loop, and R_{d1} is the SNR in the data channel. The above results indicate the importance of establishing the proper SNR in the carrier tracking loop. If this is not done, a significant loss over the theoretical performance is quite pronounced.

b. Losses in two-way links. For two-way systems, the probability that the data detector will err in its decision

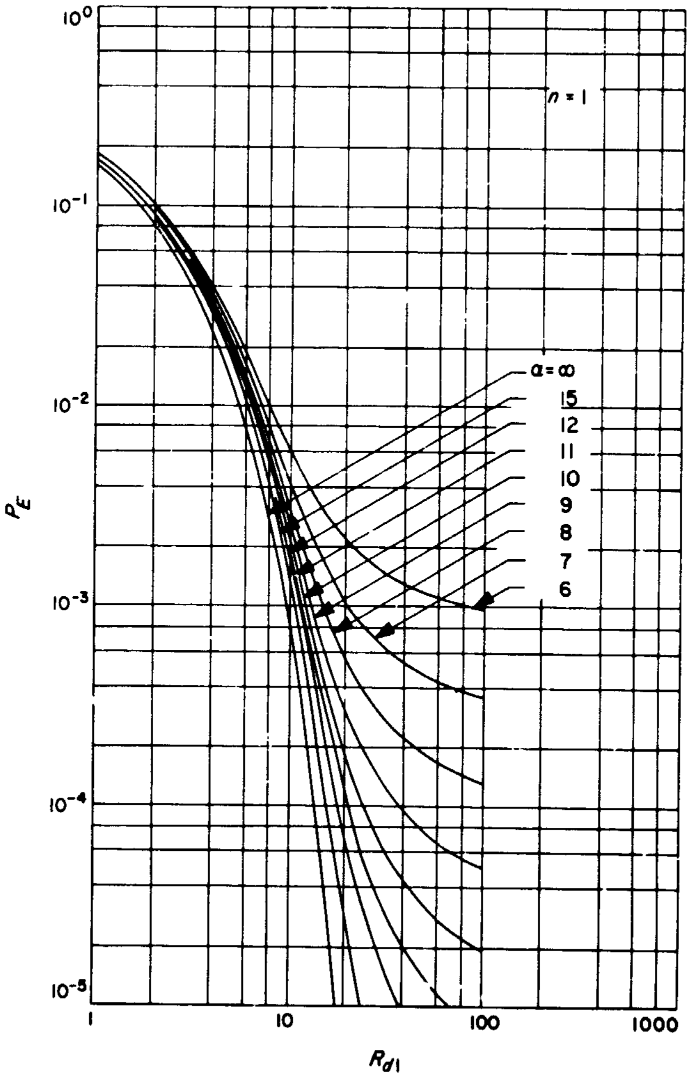

Fig. 13. Error probability vs the SNR R_{d1} for various values of the parameter α

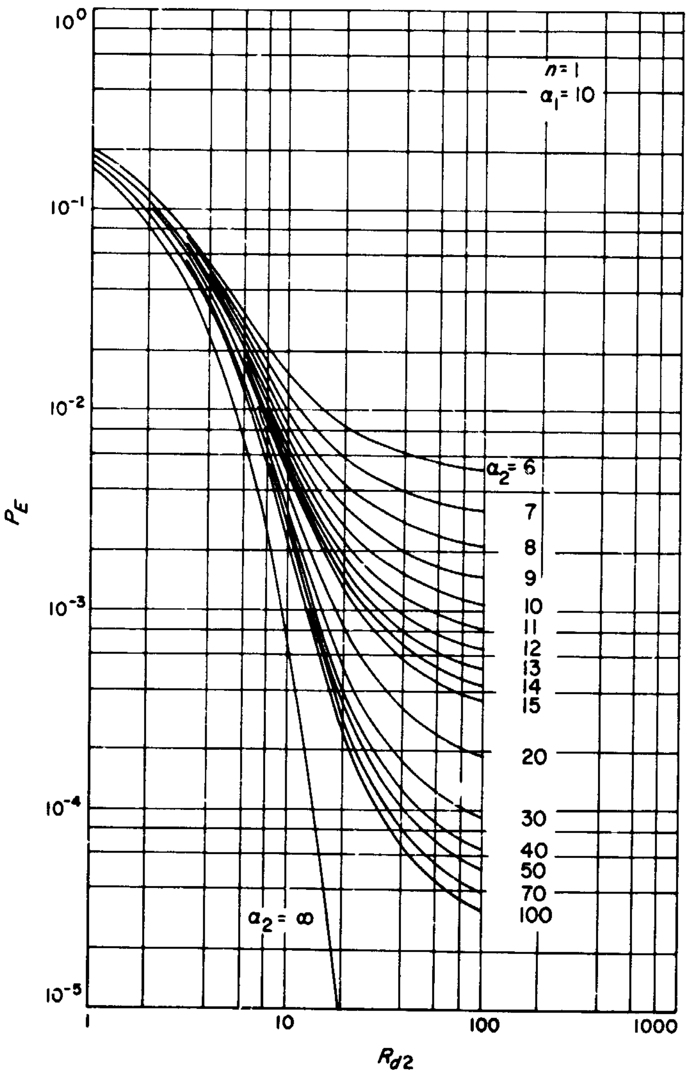

Fig. 14. Error probability vs the SNR R_{d2} for various values of the parameter α_1 with $\alpha_2 = 10$

may be evaluated following the procedure given in Ref. 14. In fact, it is easy to show that the average bit error probability is given by, Ref. 14,

$$P_E(2) = \int_{-\pi}^{\pi} p(\phi_2)\,\mathrm{Erfc}\,[((1 - m_2^2)\,R_2)^{1/2} \cos\phi_2]\,d\phi_2 \tag{20}$$

This equation has been plotted in Figs. 15–17 for various values of the parameters α_1 and α_2, where $(1 - m_2^2)\,R_2 = R_{d2}$ is the SNR existing in the data channel. These figures show the effect of varying α_1 (which, in practice, is approximately equal to the SNR existing in the vehicle's carrier tracking loop) when the SNR in the ground receiver's carrier tracking loop α_2 is held constant. Thus, the losses due to noisy carrier references are clearly manifested. These figures also indicate that the selection of the modulation factors m_1 and m_2 must take these into consideration. This selection is the subject of the next subsection.

5. Power Allocation and System Performance

In this subsection we treat the problem of dividing the power between the carrier component and the sidebands so as to minimize the probability of error. Simple formulas will be developed which allow the design engineer to compute the modulation factors m_n, $n = 1, 2$ without the aid of a digital computer. Finally, design curves will be given which allow one to make engineering tradeoffs and carry out the particular design.

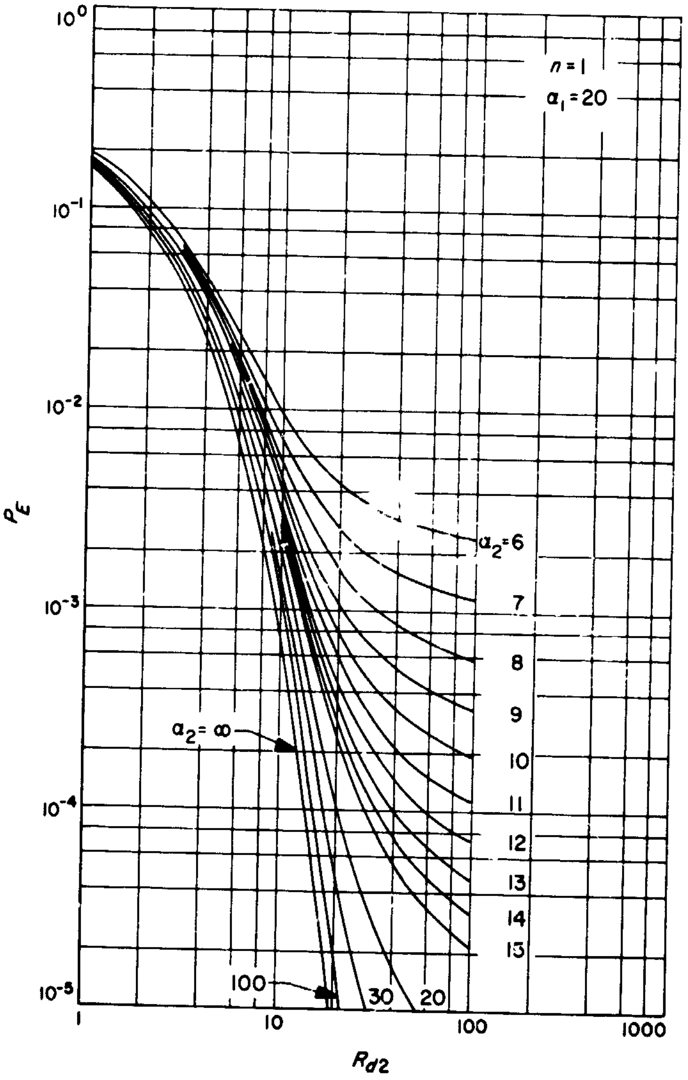

Fig. 15. Error probability vs the SNR R_{d2} for various values of the parameter α_1 with $\alpha_2 = 20$

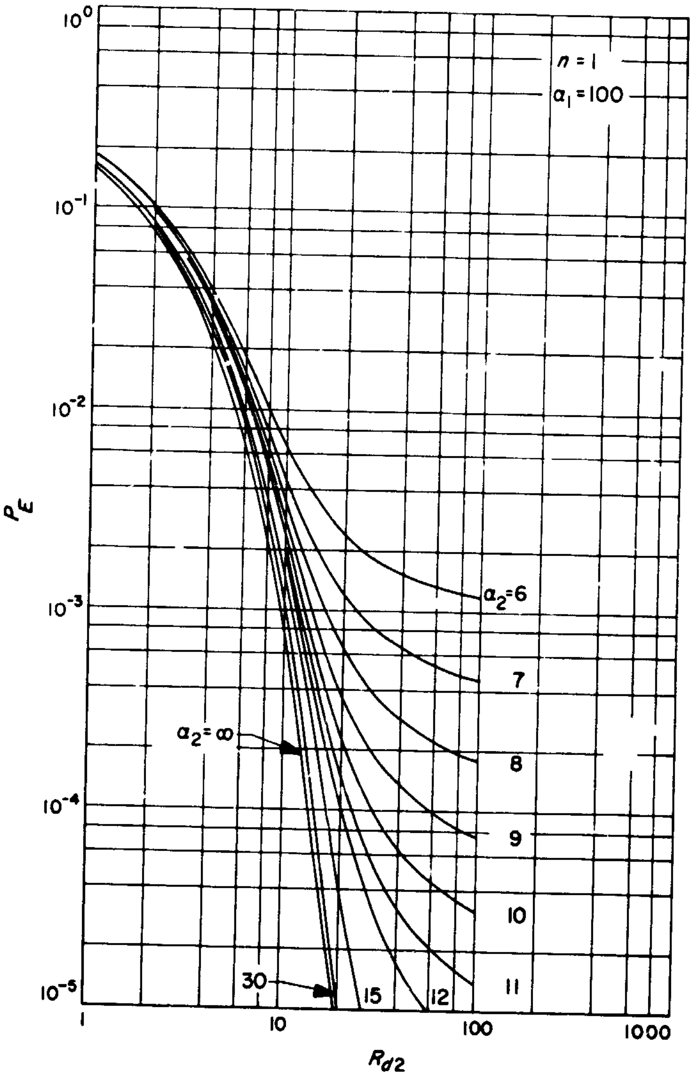

Fig. 16. Error probability vs the SNR R_{d2} for various values of the parameter α_1 with $\alpha_2 = 100$

a. Allocation of power in two-way systems. From Eq. (20) it is apparent that any attempt to find the value of m_2, which minimizes $P_E(2)$ by the method of differentiation, immediately presents formidable difficulties; however, the surface generated by Eq. (20) has been studied on the IBM 7090 computer. The procedure used by the machine was to search for that value of m_2 which minimizes $P_E(2)$ and then evaluate $P_E(2)$. The results of these computations are illustrated in Fig. 17. This figure is a plot of the optimum value of m_2^2, say m_{20}^2, versus R_2 for various values of δ_2 with $\alpha_1 = 9\,\text{db}$. This value of α_1 corresponds to a near threshold condition in the spacecraft's carrier tracking loop. Fig. 18 represents a plot of the system error probability versus R_2, for various values of δ_2, when the power is optimally divided between the carrier and sidebands. Notice from this figure that the $P_E(2)$ versus R_2 characteristic exhibits a bottoming behavior, i.e., the system exhibits an irreducible error probability. This behavior is due to the presence of additive noise on the up-link and may be eliminated by using a clean carrier reference in the vehicle or by increasing the up-link SNR to a point where the phase-jitter in the vehicle's carrier tracking loop becomes negligible.

The irreducible error probability, say $P_{E_{ir}}(2)$, may be obtained from Eq. (20) by letting N_{02} approach zero, i.e., α_2 approaches infinity, and R_2 approaches infinity. Taking the limit, we find that the integration is zero in the interval where $\cos\phi_2 \geqq 0$ and becomes

$$P_{E_{ir}}(2) = \int_{\pi/2}^{\pi} \frac{\exp(\alpha_1 \cos\phi_2)}{\pi I_0(\alpha_1)} d\phi_2 \qquad (21)$$

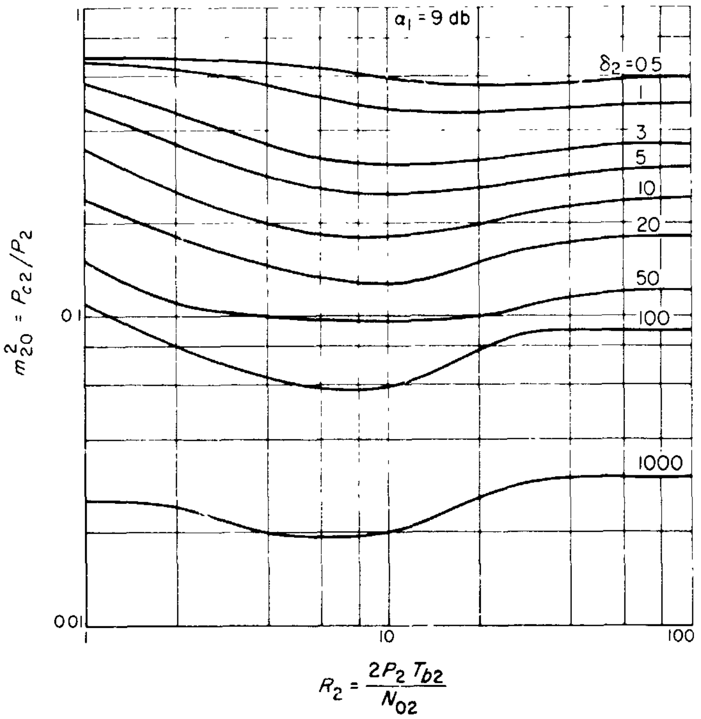

Fig. 17. Optimum modulation factor m_{20}^2 vs R_2 for various values of the parameter δ_2 with $\alpha_2 = 9$ db

where $\cos\phi_2 < 0$. This function has been studied in SPS 37-37, Vol. IV, p. 287.

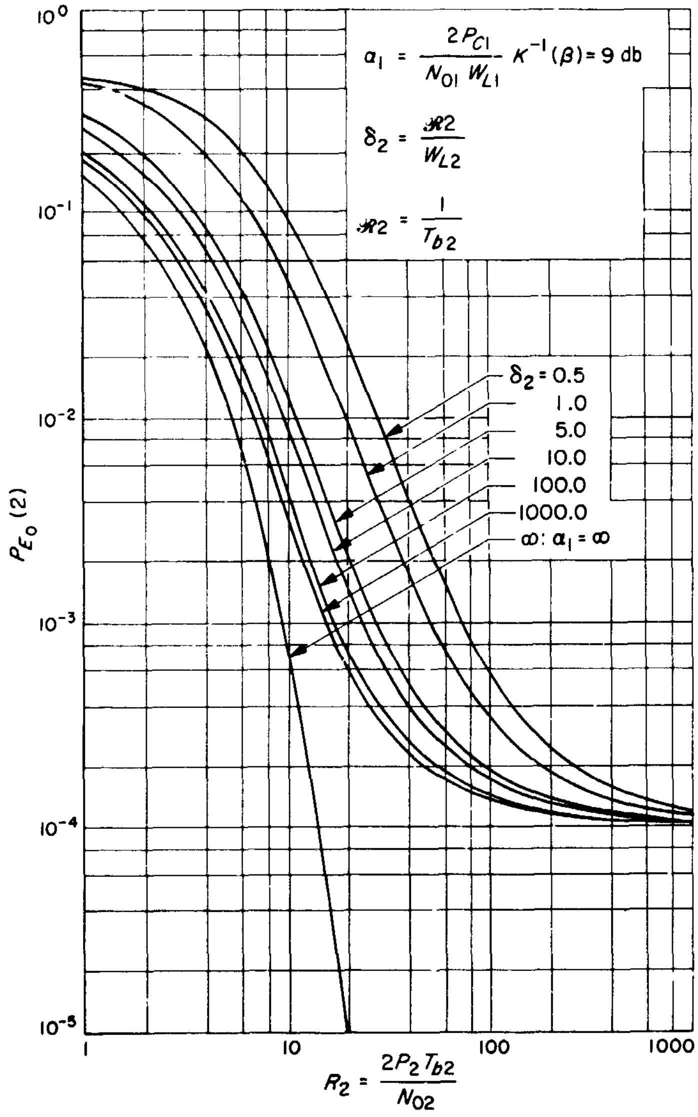

Fig. 18. Minimum error probability vs R_2 for various values δ_2 with $\alpha_1 = 9$ db

In practice, it is convenient to have an approximate formula which specifies the optimum, say m_{20}^2, value of the modulation factor as a function of the various system parameters. This approximate formula is easily obtained by assuming that the SNR existing in the individual carrier tracking loops is large enough that the linear PLL theory applies. In this case, the probability density $p(\phi)_2$ of the phase-error becomes Gaussian with variance $\sigma_{\phi_2}^2 = \alpha_1^{-1} + \alpha_2^{-1}$ and the system error probability of Eq. (20) reduces to

$$P_E(2) = \frac{1}{(2\pi\sigma_{\phi_2}^2)^{1/2}} \int_{-\infty}^{\infty} \exp\left[-\frac{\phi_2^2}{2\sigma_{\phi_2}^2}\right] \text{Erfc}\,[((1 - m_2^2)\,R_2)^{1/2} \cos\phi_2]\, d\phi_2 \tag{22}$$

Differentiating Eq. (22) with respect to m_2 and equating the result to zero yields

$$E\left[\exp\left\{-\frac{R_2(1 - m_2^2)}{2}\cos^2\left(\frac{y}{(\alpha_1^{-1} + m_2^2\delta_2 R_2)^{1/2}}\right)\right\}\right.$$
$$\cdot \frac{m_2\delta_2 R_2(1 - m_2^2)^{1/2}}{(\alpha_1^{-1} + m_2^2\delta_2 R_2)^{3/2}}\, y \sin\left(\frac{y}{(\alpha_1^{-1} + m_2^2\delta_2 R_2)^{1/2}}\right)$$
$$\left. - \left(\frac{m_2^2}{1 - m_2^2}\right)^{1/2} \cos\left(\frac{y}{(\alpha_1^{-1} + m_2^2\delta_2 R_2)^{1/2}}\right)\right] = 0 \tag{23}$$

where $E(\cdot)$ denotes the mathematical expectation of the quantity in the parenthesis, and y is a random variable which is normal (0, 1). Carrying out this expectation and solving for the value of m_2^2 which produces the minimum, gives

$$m_{20}^2 = \frac{\delta_2(R_2 - 1) - \alpha_1^{-1}(1 + 2\delta_2)}{2\delta_2 R_2 (1 + \delta_2)}$$

$$+ \frac{\{[\delta_2(R_2 - 1) - \alpha_1^{-1}(1 + 2\delta_2)]^2 + 4\delta_2(1 + \delta_2)[\delta_2 R_2 + \alpha_1^{-1}(R_2 - \alpha_1^{-1})]\}^{1/2}}{2\delta_2 R_2 (1 + \delta_2)} \tag{24}$$

This approximate formula, of course, is good in the region where the linear PLL theory applies.

b. Power allocation in one-way systems. The mathematical model, which we have established for two-way systems, reduces to the mathematical model for one-way systems if one allows B_{L2} to approach zero and replaces all subscripts on parameters which possess a two by one. In this case, the expression for the average error probability becomes

$$P_E(1) = \int_{-\pi}^{\pi} p(\phi_1)\,\mathrm{Erfc}\,[((1 - m_1^2)R_1)^{1/2}\cos\phi_1]\,d\phi_1 \tag{25}$$

where $p(\phi_1)$ is given by

$$p(\phi_1) = \frac{\exp[m_1^2\delta_1 R_1 \cos\phi_1]}{2\pi I_0(m_1^2\delta_1 R_1)}; \qquad |\phi_1| \leq \pi \tag{26}$$

and

$$\delta_1 = \frac{\mathcal{R}_1}{w_{L1}} = \frac{1}{T_{b1}w_{L1}}; \qquad R_1 = \frac{2P_1T_{b1}}{N_{01}} \tag{27}$$

As before, the value of m_1^2 which minimizes $P_E(1)$ in Eq. (25) has been found by use of the IBM 7090 computer. This value, say m_{10}^2, is plotted in Fig. 19 versus R_1 for various values of δ_1. The error probability corresponding to this minimum, say $P_{E_0}(1)$, is illustrated in Fig. 20 versus R_1 for various values of δ_1.

Again an approximate formula which specifies the optimum m_{10}^2 as a function of the various system parameters is of interest to the design engineer. This approximate formula may be obtained from Eq. (24) by letting α_1 approach infinity and replacing all parameters with subscripts two by ones. Thus, from Eq. (24) we have with $\alpha_1 = \infty$,

$$m_{10}^2 = \frac{(R_1 - 1) + ((R_1 - 1)^2 + 4R_1(1 + \delta_1))^{1/2}}{2R_1(1 + \delta_1)} \tag{28}$$

This approximate formula is plotted in Fig. 21 versus R_1 for various values of δ_1 and may be used to compare with the exact results given in Fig. 16.

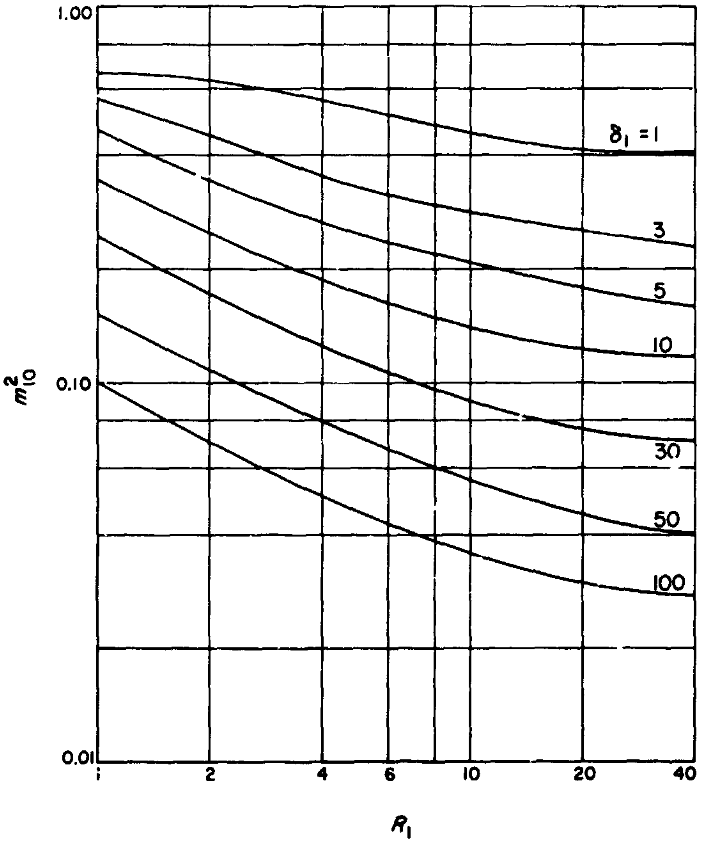

Fig. 19. Optimum modulation factor m_{10}^2 vs R_1 for various values of the parameter δ_1

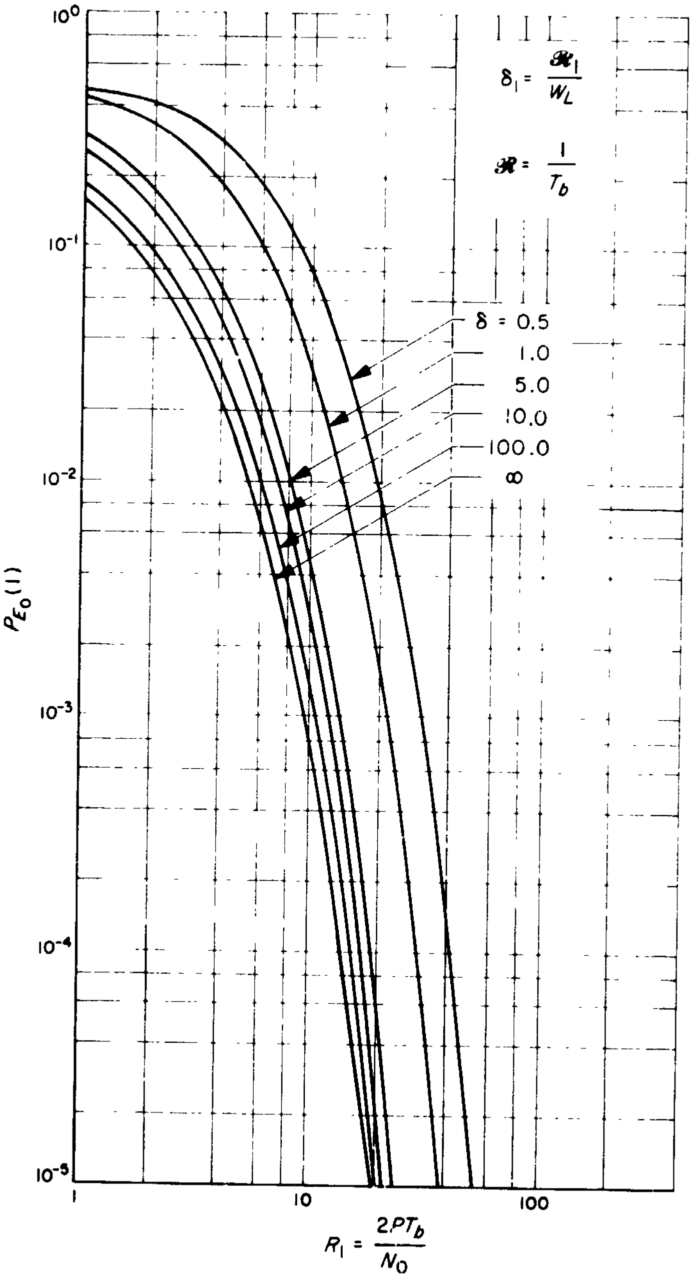

Fig. 20. Minimum error probability vs R_1 for various values of the parameter δ_1

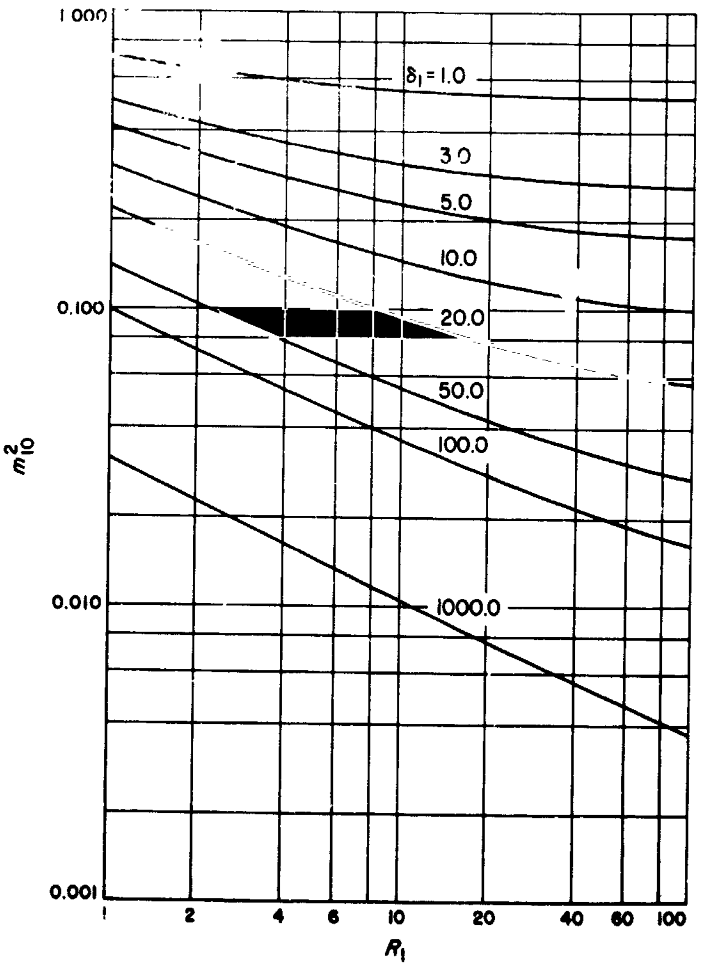

Fig. 21. Optimum modulation factor m_{10}^2 vs R_1 for various values of the parameter δ_1 (approximate formula used)

C. Combinatorial Communications: On The Number Of Information Bits In Certain Cyclic Codes, *R. McEliece*

1. Introduction

Any binary group code can be described by its generator matrix whose rows may be thought of as a basis for the vector space which constitutes the code. It is always important to know the number of information bits in such a code. However, if the code is defined only by its generator matrix, determining the number of bits is difficult, because it is equivalent to finding the rank of the generator matrix. However, when the code is cyclic, it frequently happens that the generator matrix is what is called a *circulant*, and in this case it is possible to greatly simplify the task of finding the number of information bits. In subsection 2 an algebraic method for calculating the rank of a circulant will be developed, and in subsection 3 an application to the so-called quadratic residue codes will be given.

2. Algebraic Theory

Definition: An $n \times n$ matrix with entries in a field F which has the form

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \cdots & a_n \\ a_n & a_1 & a_2 \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 \cdots & a_{n-2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_2 & a_3 & a_4 \cdots & a_1 \end{bmatrix}$$

is called a *circulant* matrix.

For our applications, it will be necessary to compute the rank of a circulant matrix. The following theorem is useful:

Theorem 1: Let $F[x]$ be the ring of polynomials with coefficients from the field F. Define the polynomial $A(x) = a_1 + a_2x + a_3x^2 + \cdots + a_nx^{n-1}$. Let $r(x) = \gcd(x^n - 1, A(x))$ in $F[x]$. Then the rank of A is equal to n-degree $(r(x))$, where A is the circulant whose first row is $(a_1, a_2, \cdots, a_n)$.

Proof: In $F[x]$ let I be the principal ideal generated by the polynomial $x^n - 1$, and denote by $F_n[x]$ the quotient ring $F[x]/I$. Since $F[x]$ is a principal ideal domain, so is $F_n[x]$. The elements in $F_n[x]$ are regarded as polynomials modulo $x^n - 1$.

Let us view $F_n[x]$ merely as a vector space of dimension n over F. Then the elements $A(x), x \cdot A(x), \cdots, x^{n-1}A(x)$ generate a subspace $\mathcal{A}$. Furthermore, the dimension of $\mathcal{A}$ is precisely the rank of the matrix A. But $\mathcal{A}$ is actually an ideal in the ring $F_n[x]$. This is because the given basis for $\mathcal{A}$ is only permuted when multiplied by the element x, and x generates the whole ring $F_n[x]$. Hence $\mathcal{A}$ is the principal ideal $(A(x))$. We now need the following:

Lemma 1: If $P \epsilon F_n[x]$, then $(P) = (\gcd(P, x^n - 1))$.

Proof: We return temporarily to $F[x]$, where the ideal structure is clearer. Two polynomials P and Q generate principal ideals (P) and (Q), and $(P) \supseteq (Q)$ if and only if P divides Q. We suppose that Q is the kernel of a ring homomorphism θ: $F[x] \rightarrow F[x]/(Q)$. Then the usual sort of calculation shows that $\theta(P) = P \cup Q/Q$. But in $F[x]$, $P \cup Q$ is principal and is generated by gcd (P, Q). Our lemma is the case $Q = x^n - 1$.

Lemma 1 allows us to complete the proof of Theorem 1, for it shows that $(A(x)) = (\gcd(A(x), x^n - 1))$. Suppose $\gcd(A(x), x^n - 1) = B(x)$. Then $B(x)$ divides $x^n - 1$. It is clear that the codimension of (B) in $F_n(x)$ is the degree d of B, since $1, x, \cdots, x^{d-1}$ are linearly independent (mod B), but there can be no larger set. Thus, the dimension of (B) is $n - d$, as asserted. Theorem 1 is proved.

3. An Application

If p is a prime, the quadratic residue codes we shall be interested in can be described by their generator matrices Q_p, which are circulants over $GF[2]$. These matrices are, therefore, completely described by their first rows $(a_0, a_1, \cdots, a_{p-1})$. Using the familiar Legendre symbol (a/p), we describe the first row:

$$a_0 = 0$$

$$a_k = \begin{cases} 1 & \text{if} \left(\dfrac{k}{p}\right) = +1 \\ 0 & \text{if} \left(\dfrac{k}{p}\right) = -1 \end{cases}$$

The code is a cyclic linear code with block length p, and the number of information symbols is the rank of the matrix Q_p. Since the assignment $a_0 = 0$ was arbitrary, we shall be interested in not only the rank of Q, but also Q', which is also a circulant but has $a_0 = 1$. We now define the polynomials

$$Q(x) = Q_p(x) = \sum_{k=0}^{p-1} a_k x^k$$

In view of the lemma of subsection 2, we shall be interested in finding the degree of the polynomials $(Q(x), x^p + 1)$ and $(Q(x) + 1, x^p + 1)$ for various odd primes p.

Let $P(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$. Then $x^p + 1 = (x + 1) P(x)$ over $GF(2)$. It is easy to determine whether or not $x + 1$ is a factor of $Q_p(x)$; if the number of nonzero coefficients is even, it is; if the number is odd, it is not. It is, therefore, sufficient to compute the degree of $(Q(x), P(x))$ and $(Q(x) + 1, P(x))$. As a final preliminary, set $N(x) = P(x) + Q(x) + 1$; i.e., $N(x)$ indicates the *nonresidues* of p.

Lemma 2: Let a be a residue (mod p). Then (mod $x^p + 1$)

$$Q(x^a) = \begin{cases} Q(x) & \text{if} \left(\dfrac{a}{p}\right) = +1 \\ N(x) & \text{if} \left(\dfrac{a}{p}\right) = -1 \end{cases}$$

$$N(x^a) = \begin{cases} N(x) & \text{if} \left(\dfrac{a}{p}\right) = +1 \\ Q(x) & \text{if} \left(\dfrac{a}{p}\right) = -1 \end{cases}$$

Proof: The polynomial

$$Q(x^a) \,(\text{mod}\, x^p + 1) = \sum_{(k/p)=+1} x^{ka\,(\text{mod}\, p)}$$

But as k runs through the quadratic residues, ka either runs through the quadratic residues or through the nonresidues, according to whether $(a/p) = +1$ or -1. The proof for $N(x)$ is the same.

We are now in a position to compute the degrees of $(Q(x), P(x))$ and $(Q(x) + 1, P(x))$, but we must first distinguish between two cases.

Case I: $(2/p) = -1$; i.e., $p \equiv \pm 3 \pmod 8$. By Lemma 2, $N^2(x) = N(x^2) \equiv Q(x) \pmod{x^p + 1}$, and $Q^2(x) \equiv N(x)$ × (mod $x^p + 1$). We may replace the modulus $x^p + 1$ by $P(x)$ in both cases, and obtain:

$$\begin{aligned} (Q(x), P(x)) &= (Q(x), Q(x) + N(x) + 1) \\ &= (Q(x), N(x) + 1) \\ &= (Q(x), Q^2(x) + 1) \\ &= \text{constant} \end{aligned}$$

$$\begin{aligned} (Q(x) + 1, P(x)) &= (Q(x) + 1, Q(x) + N(x) + 1) \\ &= (Q(x) + 1, N(x)) \\ &= (Q(x) + 1, Q^2(x)) \\ &= \text{constant} \end{aligned}$$

Here in both cases the degree is zero, and so the rank of the matrix Q is either p or $p - 1$, depending upon whether $p \equiv +1$ or $-1 \pmod 4$. (The rank of Q' is $p - 1$ or p.)

Case II: $(2/p) = +1$; i.e., $p \equiv \pm 1 \pmod 8$. By Lemma 2, $Q^2(x) \equiv Q(x) \pmod{x^p + 1}$. Consequently, if F is a splitting field for $x^p + 1$, we see that $Q(\delta) = 0$ or 1 for all $\delta \in F$. Thus, each root of $P(x)$ is either a root of $Q(x)$ or of $Q(x) + 1$. Therefore,

$$\begin{aligned} P(x) &= (P(x), Q(x))\,(P(x), Q(x) + 1) \quad (\text{mod}\, x^p + 1) \\ &= (P(x), Q(x))\,(P(x), N(x)) \qquad\quad (\text{mod}\, x^p + 1) \end{aligned}$$

Now let e be any nonresidue of p, and suppose that δ is a pth root of 1 in F with $Q(\delta) = 0$. Then by Lemma 2, $N(\delta^e) = Q(\delta) = 0$. Similarly, $N(\delta) = 0$ implies $Q(\delta^e) = 0$. Thus, there are as many pth roots of 1 with $Q(\delta) = 0$ as there are with $Q(\delta) = 1$. Hence

$$\text{degree}\,(P(x), Q(x)) = \text{degree}\,(P(x), Q(x) + 1) = \frac{p-1}{2}$$

So in this case the rank of the matrix Q is either $(p + 1)/2$ or $(p - 1)/2$, according as $p \equiv 1 \pmod 4$ or $p \equiv -1 \pmod 4$. The rank of Q' is either $(p - 1)/2$ or $(p + 1)/2$ in this case.

In both of these cases, the codes corresponding to the higher dimensions (p and $(p + 1)/2$) may be obtained from those of lower dimension ($p - 1$ and $(p - 1)/2$) by the addition of the all-ones vector to the code word. We state these results in a theorem:

Theorem 2: Let p be an odd prime. We define a *quadratic residue code* of length p in terms of its generator matrix Q, which is a circulant with first row $(a_0, a_1, \cdots, a_{p-1})$, where

$$a_0 = \begin{cases} 1 & \text{if } p \equiv -1 \pmod 4 \\ 0 & \text{if } p \equiv +1 \pmod 4 \end{cases}$$

$$a_k = \begin{cases} 1 & \text{if } \left(\dfrac{k}{p}\right) = +1 \\ 0 & \text{if } \left(\dfrac{k}{p}\right) = -1 \end{cases} \qquad k = 1, 2, \cdots, p-1$$

Then the number of information bits in the code is $p - 1$ if $p \equiv \pm 3 \pmod 8$ and $(p-1)/2$ if $p \equiv \pm 1 \pmod 8$. We may increase the number of bits by 1 in each case by adjoining the all - 1's vector to the code.

4. Conclusion

In subsection III we applied Theorem 1 to the quadratic residue codes. It should be emphasized, however, that while it may not often happen that we will be able to calculate code dimensions in a theoretical way, Theorem 1 is still of use. For Theorem 1 reduces the difficult general problem of calculating the dimension of a cyclic code to the computationally simpler problem of determining the greatest common divisor of two polynomials. This is, of course, easily done by means of Euclid's algorithm, which is very efficient and is very easy to program on a digital computer.

D. Combinatorial Communications: A Rational Algorithm for Marsh's Cubic Transformation, *S. W. Golomb*[6]

1. Introduction

Given an irreducible polynomial $f(x)$ of degree n over $GF(2)$, it is frequently desired to generate others of the same type. The two transformations

$$T: f(x) \rightarrow f(x+1)$$

and

$$U: f(x) \rightarrow x^n f(1/x)$$

have the property that both irreducibility and degree are preserved. Also, the relationship of the roots before and after is clear. ($T: \alpha \rightarrow \alpha + 1$ and $U: \alpha \rightarrow \alpha^{-1}$.) However, T and U together generate a group of only six transformations, which thus severely limit the number of new polynomials obtainable from a given one in this way.

In Ref. 19, March introduces the "cubic transformation"

$$M: f(x) \rightarrow f(x^{1/3}) f(\omega x^{1/3}) f(\omega^2 x^{1/3}) = f^*(x)$$

where $\omega^3 = 1$. It is easily seen that the roots of $f^*(x)$ are the cubes of the roots of $f(x)$. In particular, for odd degree n, $2^n - 1$ is not divisible by 3, and the transformation M preserves not only irreducibility but also the degree of primitivity of the roots. In a variety of cases, iteration of M enables one to generate *all* irreducible polynomials of degree n from a given one. (These degrees include $n = 3, 5, 7, 13, 17,$ and 19.)

The purpose of this note is to describe a "rational" algorithm for effecting the transformation M, which is useful in preparing tables[7] (Ref. 20). The procedure is rational in the sense that ω and ω^2 do not appear in the final result $f^*(x)$ nor in the intermediate computations.

2. The Algorithm

Divide the exponents of the terms in $f(x)$ into three classes: A, B, and C, according to the residue class of the exponent modulo 3. We produce the set of exponents for $f^*(x)$ from those for $f(x)$ by the following 3 steps:

(1) Copy the exponents of $f(x)$.

(2) Adjoin all numbers $(2u_1 + u_2)/3$ where u_1 and u_2 are distinct exponents of $f(x)$ in the same residue class modulo 3.

(3) Adjoin all numbers $(a + b + c)/3$ where $a \in A$, $b \in B$, and $c \in C$.

Any exponent for $f^*(x)$ which is produced an *even* number of times by these operations must be discarded; if produced an *odd* number of times, it should be retained

[6] Prepared under contract 951076 with the University of Southern California, Department of Electrical Engineering, Los Angeles, California.

[7] S. W. Golomb, L. R. Welch, and A. W. Hale's memorandum *On the Factorization of Trinomials Over GF(2)*, Jet Propulsion Laboratory, Pasadena, California.

(once). If any of the three categories A, B. C is empty, then step (3) is vacuous. If a category has less than two members, it does not contribute to step (2).

Example 1. Let $f(x) = x^5 + x^2 + 1$. Then the categories are

A	B	C
0		2, 5

To produce $f^*(x)$, we follow the three steps:

	A	B	C	
step 1	0		2, 5	copy
step 2	3	4		$\frac{2\times 2+5}{3}, \frac{2\times 5+2}{3}$
step 3				vacuous
mod 2 sum	0, 3	4	2, 5	

Thus $f^*(x)$ has the exponents 0, 2, 3, 4, 5 and

$$f^*(x) = x^5 + x^4 + x^3 + x^2 + 1 .$$

Example 2. We iterate the transformation, this time starting with $f(x) = x^5 + x^4 + x^3 + x^2 + 1$. To form $f^{**}(x)$, we follow the three steps:

	A	B	C	
step 1	0, 3	4	2, 5	copy
step 2	3	1, 4	2	$\frac{2\cdot 0+3}{3}, \frac{2\cdot 3+0}{3}, \frac{2\cdot 2+5}{3}, \frac{2\cdot 5+2}{3}$
step 3	3, 3	4	2	$\frac{0+4+2}{3}, \frac{0+4+5}{3}, \frac{3+4+2}{3}, \frac{3+4+5}{3}$
mod 2 sum	0	1, 4	2, 5	

Thus, $f^{**}(x) = x^5 + x^4 + x^2 + x + 1$.

The reader is invited to verify $f^{***}(x) = x^5 + x^4 + 1$.

3. Proof of Algorithm

We wish to show that $f^*(y^3) = f(y)f(\omega y)f(\omega^2 y)$ can be obtained in the manner just described, where $y = x^{1/3}$. Write $f(y) = f_0(y) + f_1(y) + f_2(y)$, where $f_i(y)$ contains precisely those terms of $f(y)$ with the exponent congruent to i modulo 3. Then

$$f(\omega y) = f_0(y) + \omega f_1(y) + \omega^2 f_2(y)$$

and

$$f(\omega^2 y) = f_0(y) + \omega^2 f_1(y) + \omega f_2(y)$$

Thus,

$$\begin{aligned} f^*(y^3) &= (f_0 + f_1 + f_2)(f_0 + \omega f_1 + \omega^2 f_2)(f_0 + \omega^2 f_1 + \omega f_2) \\ &= (f_0^3 + f_1^3 + f_2^3) + (1 + \omega + \omega^2)(f_0 f_1^2 + f_1 f_2^2 + f_2 f_0^2 \\ &\quad + f_0^2 f_1 + f_1^2 f_2 + f_2^2 f_0) + \begin{vmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{vmatrix} f_0 f_1 f_2 \\ &= (f_0^3 + f_1^3 + f_2^3) + f_0 f_1 f_2 \end{aligned}$$

since $1 + \omega + \omega^2 = 0$, while the determinant is 1. (The matrix is nonsingular by linear independence of the rows, and the determinant is rational by symmetry in ω and ω^2.)

The exponents in f_0, f_1, and f_2 are those in the classes A, B, and C, respectively. The exponents of f_1^3 are: (1) the triples of the exponents of f_1, and (2) the sums $2u_1 + u_2$, where u_1 and u_2 are distinct exponents of f_1. The exponents of $f_0 f_1 f_2$ are: (3) all sums of the form $a + b + c$, with $a \epsilon A$, $b \epsilon B$, and $c \epsilon C$. Allowing for $y^3 = x$, these are the three steps in the algorithm for finding $f^*(x)$ from $f(x)$.

E. Combinatorial Communications: On Enumerative Equivalence of Group Elements, *S. W. Golomb*[a] *and A. W. Hales*[b]

1. Introduction

Let G be a finite group (of order $|G|$) operating on a finite set S. By a well-known formula of Burnside (Ref. 21), widely exploited since the appearance of Polya's paper (Ref. 22), the number C of *orbits* (equivalence classes of elements in S under the operations of G) is given by

$$C = \frac{1}{|G|} \sum_{g \epsilon G} I(g) \qquad (1)$$

where $I(g)$ is the number of fixed points of S under the group element g. It is well known (Ref. 23) that, considering S as a representation of G, $I(g)$ is a *group character*. Thus, if g_1 and g_2 are conjugate elements of G, $I(g_1) = I(g_2)$ for *all* representations S.

The purpose of this article is to characterize the circumstances under which two group elements, g_1 and g_2, of the abstract group G satisfy $I(g_1) = I(g_2)$ for every representation S of G. Two group elements with this property will be called *(weakly) enumeratively equivalent* (this is clearly an equivalence relation on G), and in applying Eq (1) it suffices to compute $I(g)$ only once for each enumerative equivalence class in G.

We define two group elements, g_1 and g_2, of the abstract group G to be *strongly enumeratively equivalent* if they have the *same set* of fixed points, no matter what representation of G is considered. We call two group elements, g_1 and g_2, *related* (and say g_1 *is a relative of* g_2) if they generate the same cyclic subgroup of G, i.e., if each is a power of the other. The characterization theorems for strong weak enumerative equivalence are as follows:

Theorem 1. Two elements, g_1 and g_2, of an abstract group G are strongly enumeratively equivalent if and only if g_1 is a relative of g_2.

Theorem 2. Two elements, g_1 and g_2, of an abstract finite group G are weakly enumeratively equivalent if and only if g_1 is a conjugate of a relative of g_2.

Theorem 2 is no longer true if the word "finite" is omitted. We shall indicate the appropriate modifications in this case. Finally, we shall discuss those groups G in which "g_1 a relative of g_2" implies "g_1 a conjugate of g_2," and vice versa.

2. An Example

In Fig. 22, the points a, b, c are the vertices of an equilateral triangle, and the line df is perpendicular to the triangle, with e as the midpoint of both the segment df and the triangle abc. On the set $S = \{a, b, c, d, e, f\}$, we have a group of operators $G = \{A, B, C, D, E, F\}$, where A, B, and C are 180-deg rotations of S around the indicated medians of the triangle, D and F are the ±120-deg

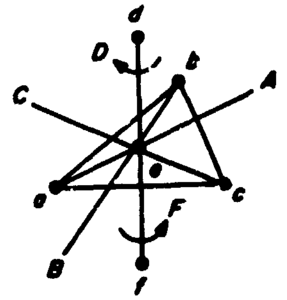

Fig. 22. The dihedral group

[a] Prepared under JPL contract 951076 with the University of Southern California, Department of Electrical Engineering, Los Angeles, California.

[b] Consultant from the University of California, Los Angeles, California.

rotations of S about the perpendicular line df, and E is the identity operator. (Abstractly, G is the same as the dihedral group of the equilateral triangle and the symmetric group on three letters.) The fixed points are listed and counted in Table 2.

Table 2. Fixed points

Group element	Fixed points	I (g)
E	a, b, c, d, e, f	6
A	a, e	2
B	b, e	2
C	c, e	2
D	d, e, f	3
F	d, e, f	3

Formula (1) says that $1/6 \Sigma I(g)$, which is 3, should equal the number of *orbits*. It is clear that the orbits are $\{e\}$, $\{a,b,c\}$, and $\{d,f\}$. Since D and F are *related*, by Theorem 1 they must have the same fixed points (viz., d, e, and f). The operators A, B, and C are not related and have different sets of fixed points. However, since they are conjugate, by Theorem 2 they all have the same $I(g)$.

By Theorems 1 and 2 respectively, the strong enumerative equivalence classes in G are $\{E\}$, $\{A\}$, $\{B\}$, $\{C\}$, $\{D,F\}$ and the weak enumerative equivalence classes are $\{E\}$, $\{A,B,C\}$, and $\{D,F\}$. The example verifies that for this group, the classes are certainly no larger than allowed by the theorems.

3. Strong Enumerative Equivalence

Lemma 1. If g_1 is a power of g_2, then every fixed point of g_2 is also a fixed point of g_1.

Proof: Let s_0 be any fixed point of g_2, so that $g_2(s_0) = s_0$, and let $g_1 = g_2^a$. Then

$$g_1(s_0) = g_2^a(s_0) = g_2 g_2 \cdots g_2(s_0) = s_0 .$$

Lemma 2. If g_1 and g_2 are related elements of the group G, then they have the same set of fixed points in any representation S of G.

Proof: g_1 and g_2 are related if and only if each is a power of the other. Then by Lemma 1, if they are related, they have the same fixed points.

Lemma 3. If g_1 and g_2 are nonrelated elements of G, there exists a representation S in which g_1 and g_2 do not have the same set of fixed points.

Proof: Construct S as follows: Let all the cyclic subgroups of G, and all their left cosets, be the *elements* of S. Let G operate on S by left multiplication. (The effect of a group element is to permute a subgroup and its cosets. Hence, S is a valid representation of G.)

If g_1 and g_2 are nonrelated elements of G, they do not generate the same cyclic subgroup. Hence, there is a cyclic subgroup H containing one (say g_1) but not the other. As an element of S, H is then a fixed point of g_1 but not of g_2.

Theorem 1. Two elements, g_1 and g_2 of an abstract group G, are strongly enumeratively equivalent if and only if g_1 is a relative of g_2.

Proof: The two halves of this theorem are Lemmas 2 and 3, respectively.

4. Weak Enumerative Equivalence

We need several lemmas.

Lemma 4. If g_1 is a conjugate of a power of g_2, then $I(g_1) \geqq I(g_2)$ for every representation S of G.

Proof: By Lemma 1 of the previous section, it suffices to show that the function I is constant on conjugate classes of G. But this is proved in (Ref. 23) Theorem 3.

Lemma 5. If g_1 is not a conjugate of a power of g_2, then there is a representation S of G in which $I(g_1) < I(g_2)$.

Proof: Let H be the cyclic subgroup of G generated by g_2; let S consist of H and its left cosets; and let G operate on S by left multiplication. Then $I(g_2) \geqq 1$, because g_2 leaves H fixed. On the other hand $I(g_1) = C$, because if g_1 had a fixed point in S, say mH, we would have

$$g_1(mH) = mH$$

$$(m^{-1}g_1m) = H$$

implying that $m^{-1}g_1m$, a conjugate of g_1, is in H, and hence is a power of g_2. This contradiction proves the lemma.

Lemma 6. Two elements, g_1 and g_2, of an abstract group G are (weakly) enumeratively equivalent if and only if each is a conjugate of a power of the other.

Proof: This follows immediately from Lemmas 4 and 5.

Lemma 7. Let g_1 and g_2 be elements of finite order in a group G. Then the following are equivalent:

(1) Each of g_1, g_2 is a conjugate of a power of the other.

(2) g_1 is a conjugate of a relative of g_2.

Proof: It is clear that assertion (2) implies (1). To show that (1) implies (2), we observe that condition (1) implies both that (0 is the order function) $0(g_1) \leq 0(g_2)$ and $0(g_1) \geq 0(g_2)$. Hence, g_1 and g_2 have the same order. We know that $g_1 = x^{-1} g_2^n x$ for some x. Here g_1 and g_2^n have the same order. Hence, g_2 and g_2^n have the same order and, therefore, generate the same cyclic subgroup, i.e., g_2^n is a relative of g_2. This completes the proof.

Theorem 2. Two elements, g_1 and g_2, of an abstract finite group G are (weakly) enumeratively equivalent if and only if g_1 is a conjugate of a relative of g_2.

Proof: Since every element of a finite group is of finite order, the theorem follows from Lemmas 6 and 7.

The following example shows that the finiteness hypothesis in Lemma 7 (and hence in Theorem 2) is necessary.

Example: Let G be the split extension of $Q \times Q$ (when Q is the rationals under $+$) by an infinite cyclic group generated by x, with $x^{-1}(a, b)\,x = (2b, 2a)$. Then $(1, 0)$ is conjugate to the "power" $(0, 2)$ of $(0, 1)$, and $(0, 1)$ is conjugate to the "power" $(2, 0)$ of $(1, 0)$, but $(1, 0)$ is *not* a conjugate of a relative of $(0, 1)$.

5. Considerations of Implication

In the light of the previous results, it is interesting to ask the following two questions:

(A) For which groups G does "g_1 is a conjugate of g_2" imply "g_1 is a relative of g_2"?

(B) For which groups G does "g_1 is a relative of g_2" imply "g_1 is a conjugate of g_2"?

Question A is easy to answer.

Theorem 3. Let G be an abstract group. Then the following are equivalent:

(1) g_1 is a conjugate of g_2 implies g_1 is a relative of g_2.

(2) All subgroups of G are normal.

Proof: Assume (1). Then if H is a subgroup of G, h is in H; and g is in G. We have $g^{-1} h g$ is a relative of h and, hence, is also in H. Thus, H is normal. Conversely, suppose (2) is true. If g_1 is a conjugate of g_2, then, since the cyclic subgroup generated by g_2 is normal, g_1 must be a power of g_2. Similarly, g_2 is a power of g_1, and the proof is complete.

Groups in which every subgroup is normal are called *Hamiltonian groups*, and have been completely classified (Ref. 24). The result is this: a group is Hamiltonian if and only if it is either abelian, or the direct product of the quaternion group of order 8 by an abelian group of exponent two by an abelian group in which every element has finite odd order. This completes the answer to question A.

Question B is much more difficult to answer, and will not be dealt with here. We only point out the following facts, without proof: no group of finite odd order satisfies the condition of question B; any abelian group of exponent two does satisfy the condition; the quaternion group satisfies the condition; and all symmetric groups satisfy the condition.

F. Information Processing: Arithmetic Decoding of Cyclic Codes, II, G. Solomon

1. Summary

In SPS 37-42, Vol. IV, pp. 205–208, we introduced simplified mechanizable arithmetic decoding of maximal length shift register codes and cyclic Reed-Mueller Codes. The number of computations was proportional to the number of information bits. This article extends these procedures for more general BCH cyclic codes.

2. Main Result

Let k be even. Consider the BCH code generated via linear recursion by the polynomial

$$f(x) = (x + 1)\, f_1(x)\, f_5(x)$$

where f_1, f_5 are the irreducible polynomials of α and α^5 over GF (2), α primitive. For these codes of length $2^k - 1$, and usual dimension $2k + 1$ (only exception: $k = 4$) we have a very simple decoding procedure which will correct $[(2^k - 1)/6] - 1$ errors:

Example.

(15, 7) BCH code will correct 2 errors with this procedure.

(63, 13) BCH code will correct 10 errors with this procedure.

(255, 17) BCH code will correct 42 errors with this procedure.

This arithmetic procedure will correct, in general, slightly fewer errors than the usual BCH error correcting procedures, but is much simpler in mechanization and concept.

3. Method

Let $a = (a_i)\ i = 0, 1, \cdots, 2^k - 2$ be the received vector.

(1) Perform the permutation $\phi(a)$ by mapping $x \rightarrow x + 1$, $x \varepsilon GF(2^k)$, as in SPS 37-42, Vol. IV, pp. 205–208.

(2) Let $T\bar{a}$ be the cyclic shift of $\bar{a}$ by $(2^k - 1)/3$ stages to the right. Perform $T\phi(a)$, $T^2\phi(a)$.

(3) Compute the vector sum $\phi(a) + T\phi(a) + T^2\phi(a) = b$.

(4) Compute the weight ω of b over all coordinates but the 0th. Let $\omega' = \omega/3$.

(5) Add as mod 0 (ordinary arithmetic) to ω'.

(6) If $\omega' < (2^k - 1)/(3 \cdot 2)$, decode a_0 as 0. Otherwise, decode a_0 as 1.

(7) Cyclically shift the received vector $\bar{a}$ one stage to the right to obtain $(a_1, a_2, \cdots, a_0)$.

(8) Perform (1)–(6) on shifted vector. Decode a_1.

(9) Continue until all information bits are decoded.

4. Proof

For the codes described above, the vectors are given by $\{g(\alpha^i);\ i = 0, 1, \cdots, 2^k - 2\}$ where (SPS 37-42, Vol. IV)

$$g(x) = c_0 + Tr\,cx + Tr\,dx^5$$

α primitive, $c_0 \varepsilon GF(2)$, $c \varepsilon GF(2^k)$, $d \varepsilon GF(2^l)$ where $l|k$ and $2^l - 1$ is the order of α^5.

The coordinates $1 + \alpha^i$, $1 + \alpha^{[2^k-1/3]+i}$, $1 + \alpha^{2[(2^k-1)/3]+i} = \beta_i, \beta_{i'}, \beta_{i''}$ say, $i = 1, \cdots, (2^k - 1)/3 - 1$ are distinct in $GF(2^k)$ and take on $2^k - 1$ values; also, $(\beta_i - 1)^3 = \alpha^{3i}$, etc. It can be shown that

$$g(\beta^i) + g(\beta^{i'}) + g(\beta^{i''})$$
$$= c_0 + Tr\,c(1) + Tr\,d(1) = a_0, \qquad \text{if correct}$$

We then have $[(2^k - 1)/3] - 1$ determinations of a_0 in addition to the value of a_0 itself. Decode a_0 as 0 if the majority of these determinations is 0, and as 1, if the majority is 1. Note that, since k is even, $(2^k - 1)/3$ is an odd integer. Hence, a tie can not occur, and the decoding procedure always produces an answer regardless of how many errors were made.

One readily verifies that if at most $(2^k - 1)/6 - 1$ errors have been made, the majority decision is indeed correct.

G. Information Processing: Decoding Codes Beyond the Bose-Chaudhuri Bound,

E. R. Berlekamp[10]

It is known from combinatorial arguments that most t-error correcting Bose-Chaudhuri-Hocquenghem (BCH) codes are capable of correcting many error patterns containing more than t errors, although no feasible general algorithms for correcting correctable error patterns are known. In certain cases, decoding words with $t + e$ errors can be reduced to the solution of some simultaneous nonlinear equations in e unknowns ($e > 0$). Unfortunately, feasible methods for solving these equations are known only in a few special cases. The case of $t + 1$ errors is a particular example.

1. Introduction

In a t-error correcting binary BCH code of odd block length N, positions of the code are associated with the Nth roots of unity, which form a multiplicative subgroup of the nonzero elements of some finite field, $GF(2^k)$, of characteristic 2.

The code is constructed in such a way that various power-sum symmetric functions of the error locations are available as parity checks on the received word. In particular, the t-error correcting BCH code is chosen so that the power-sum symmetric functions $S_1, S_2, S_3, \cdots, S_{2t}$ are available, where

$$S_j = \sum_{\substack{\text{error}\\ \text{positions}}} B_i^j$$

The decoding problem is to find the S's given these S's.

In order to solve these equations, one usually proceeds in two steps. The first goal is to construct the error

[10] Consultant from the Electrical Engineering Department, University of California, Berkeley, California.

polynomial

$$\sigma(x) = \sum \sigma_i x^i = \prod_j (1 - B_j x)$$

The degree of the error polynomial is equal to the number of errors, and the roots of the error polynomial are the locations of these errors. The coefficients of the error polynomial (the σ_i) are in fact the elementary symmetric functions of the error locations. These elementary symmetric functions are related to the power-sum symmetric functions by Newton's Identities. As was shown in a previous article (Ref. 25), Newton's Identities in a field of characteristic two are conveniently expressed in generating function notation by the equation

$$S(x)\,\sigma(x) = \text{Even}(x)$$

where $S(x)$ is the generating function of the S's, given by $S(x) = 1 + S_1x + S_2x^2 + S_3x^3, \cdots$; $\sigma(x)$ is the error polynomial; and Even (x) denotes some even polynomial in x.

2. Finding $\sigma(x)$

Typically, one is given only $1 + S_1x + S_2x^2 +, \cdots, + S_{2t}x^{2t}$ (i.e., $S(x) \bmod x^{2t+1}$), and one wishes to find a $\sigma(x)$ such that $S \cdot \sigma = \text{Even} \bmod x^{2t}$. This is most readily done by an iterative process. We define a sequence of successive approximations, $\sigma^{(0)}, \sigma^{(1)}, \sigma^{(2)}, \cdots$, and an auxiliary sequence $\tau^{(0)}, \tau^{(1)}, \tau^{(2)}, \cdots$, as follows:

$$\sigma^{(0)} = \tau^{(0)} = 1$$

Then let $\Delta_1^{(n)}$ be the coefficient of x^{2n-1} in the power series of $S \cdot \sigma^{(n)}$. Now define

$$\sigma^{(n+1)} = \sigma^{(n)} - \Delta_1^{(n)} x \tau^{(n)}$$

$$\tau^{(n+1)} = \begin{cases} \dfrac{x\sigma^{(n)}}{\Delta_1^{(n)}} \text{ if } \Delta_1^{(n)} \neq 0 \text{ and if } \deg \sigma^{(n)} \leqq \deg \tau^{(n)} \\ x^2\tau^{(n)} \text{ if } \Delta_1^{(n)} = 0 \text{ or if } \deg \sigma^{(n)} > \deg \tau^{(n)} \end{cases}$$

One can readily verify that, for all n,

$$\begin{aligned} S \cdot \sigma^{(n)} &= \text{Even} \bmod x^{2n} \\ S \cdot \tau^{(n)} &= \text{Odd} + x^{2n} \bmod x^{2n+1} \\ \deg \sigma^{(\kappa)} + \deg \tau^{(n)} &= 2n \\ \sigma^{(n)}(0) &= 1 \end{aligned}$$

By further arguments, it was shown by Berlekamp (Ref. 25) that if $\rho(x)$ is any polynomial whatever $\neq \sigma^{(n)}(x)$, and if either $\deg \rho < \deg \sigma^{(n)}$ or if $\deg \sigma^{(n)} \leqq \deg \rho \leqq n$, then

$$S \cdot \rho \neq \text{Even} \bmod x^{2n}$$

Consequently, if there are no more than t errors, this iterative procedure terminates with the correct error polynomial, $\sigma^{(t)} = \sigma$.

We can express the $(n + 1)$st, $(n + 2)$nd, $\cdots$, $(n + k)$th approximations in terms of the nth approximation by the formulas

$$\sigma^{(n+k)} = \hat{f}^{(k,n)}\sigma^{(n)} + \tilde{f}^{(k,n)}\tau^{(n)}$$

$$\tau^{(n+k)} = \tilde{g}^{(k,n)}\sigma^{(n)} + \hat{g}^{(k,n)}\tau^{(n)}$$

Here $\hat{f}^{(k,n)}$ and $\tilde{f}^{(k,n)}$ represent the even and odd parts of the polynomial $f^{(k,n)}$; $\hat{g}^{(k,n)}$, and $\tilde{g}^{(k,n)}$ represent the even and odd parts of the polynomial $g^{(k,n)}$[11]. From the previous definition of the iterative algorithm, one can readily verify that the polynomials f and g must satisfy

$$f^{(0,n)} = g^{(0,n)} = 1$$

$$f^{(k+1,n)} = f^{(k,n)} - \Delta_1^{(n+k)} \times g^{(k,n)}$$

$$g^{(k+1,n)} = \begin{cases} \dfrac{xf^{(k,n)}}{\Delta_{1+k}^{(n)}} \text{ if } \Delta_1^{(n+k)} \neq 0 \text{ and } \deg \sigma^{(n+k)} \leqq \deg \tau^{(n+k)} \\ x^2 g^{(k,n)} \text{ if } \Delta_1^{(n+k)} = 0 \text{ or if } \deg \sigma^{(n+k)} > \deg \tau^{(n+k)} \end{cases}$$

In the common case where $\deg \sigma^{(n)} = \deg \tau^{(n)} = n$, we have $\deg \sigma^{(n+k)} = n + \deg f^{(k,n)}$; $\deg \tau^{(n+k)} = n + \deg g^{(k,n)}$. If $\sigma^{(n+K)} = \sigma$, then for all $k \geqq K$, $f^{(k,n)} = f^{(K,n)}$. Therefore, we can consider the polynomial $f^{(\infty,n)}$, with the obvious definition.

The branch in the iterative algorithm depends on the scalar

$$\begin{aligned} \Delta_1^{(n)} &= \left[\sum_{i=0}^{\deg(\sigma^{(n)})} S_{2n+1-i}\,\sigma_i^{(n)} \right] \\ &= S_{2n+1} - S_{2n+1}^{(n)} \end{aligned}$$

[11] Throughout this article, we let "~" and "^" denote the odd and even parts of a polynomial. Notice that "~" is the graph of an odd sin wave and "^" is the graph of an even triangular wave.

where we have set

$$S_{2n+1}^{(n)} - \left[\sum_{i=1}^{\deg(\sigma^{(n)})} S_{2n+1-i}\, \sigma_i^{(n)} \right]$$

Evidently,

$$\sigma^{(n+1)} = \sigma^{(n)} \text{ iff } \Delta_1^{(n)} = 0 \text{ iff } S_{2n+1} = S_{2n+1}^{(n)}$$

For this reason, we call $S_{2n+1}^{(n)}$ the ***anticipated value*** of S_{2n+1}. More generally, define the anticipated values of further power-sum symmetric functions by the iterative equations

$$S_0^{(n)} = 1$$

$$S_{2k}^{(n)} = (S_k^{(n)})^2$$

$$S_{2k+1}^{(n)} = \deg\left[\sum_{i=1}^{\sigma^{(n)}} S_{2k+1-i}\, \sigma_i^{(n)} \right]$$

For $k \leqq 2n$, $S_n^{(n)} = S_k$. Evidently, the polynomial

$$S^{(n)}(x) = \sum_{i=0}^{\infty} S_i^{(n)} x^i$$

is the unique solution of the equations

$$\hat{S}^{(n)}(x) = S^{(n)}(x^2)$$

$$S^{(n)}(x)\, \sigma^{(n)}(x) = \text{Even}(x)$$

For this reason, we view $S^{(n)}(x)$ as an estimate of $S(x)$, based only on $S_1, S_2, \cdots, S_{2n}$. If there are no more than n errors, then this estimate must be correct. However, if there are additional errors, then the estimated power-sum generating polynomial will differ from the true generating polynomial. Since $S^{(n)}(x) = S(x) \bmod x^{2n}$, we measure the difference by the polynomial

$$\Delta^{(n)}(x) = (S(x) - S^{(n)}(x))/x^{2n}$$

$\Delta_1^{(n)}$ is thus seen to be the first coefficient of a polynomial which represents the difference between the actual power-sum generating function and the nth approximation to it. Breaking $\Delta^{(n)}(x)$ up into even and odd parts gives

$$\Delta^{(n)} = \tilde{\Delta}^{(n)} + \hat{\Delta}^{(n)}$$

Evidently, $\Delta^{(n)} = 0 \bmod x^{2n}$. Finally, since

$$S^{(n)}(x)\, \tau^{(n)}(x) = x^{2n} \bmod x^{2n+1}$$

we define

$$A^{(n)} = \frac{S^{(n)}\, \tau^{(n)}}{x^{2n}}$$

3. Expression for $\Delta^{(n)}$

We now give an expression for $\Delta^{(n)}$ in terms of $\sigma^{(n)}$, $\tau^{(n)}$, $S^{(n)}$, $A^{(n)}$, and $f^{(\infty, n)}$. If we are given $S_1, S_2, S_3, \cdots, S_{2n}$, then we can compute $\sigma^{(n)}$, $\tau^{(n)}$, $^{n)}$, and $A^{(n)}$. The equations we are about to give will relate the two unknown polynomials $f^{(\infty, n)}$ and $\Delta^{(n)} = \Delta_1 x + \Delta_2 x^2 + , \cdots ,$. Details are omitted.

The coefficients Δ_i of x^i in $\Delta^{(n)}(x)$ are given as

$$\Delta_1 = f_1$$

$$\Delta_3 = f_1 (A_2 + f_1\tau_1 + \sigma_2 + f_2) + f_3$$

$$\begin{aligned}\Delta_5 = f_1 (A_4 &+ A_2 (f_1\tau_1 + \sigma_2 + f_2) + f_1\tau_3 + f_3\tau_1 \\ &+ \sigma_4 + f_2\sigma_2 + f_4 + (f_1\tau_1 + \sigma_2 + f_2)^2 \\ &+ f_3 (A_2 + f_1\tau_1 + \sigma_2 + f_2) + f_5\end{aligned}$$

In general, Δ_{2k-1} is given by an expression of kth degree in f_1, $(k-1)$st degree in f_2 and f_3, $(k-2)$nd degree in f_4 and f_5, $\cdots$, and 1st degree in f_{2k-2} and f_{2k-1}.

4. Decoding More than t Errors

Let us consider how these expressions can be used to decode more than t errors in a t-error-correcting Bose-Chaudhuri-Hocquenghem code. We first compute $\sigma^{(1)}$, $\tau^{(1)}$, $\sigma^{(2)}$, $\tau^{(2)}$, $\cdots$, $\sigma^{(t)}$, $\tau^{(t)}$. Then we compute the first several coefficients of $S^{(t)}$ and of $A^{(t)}$. We are then in a position to apply some of the above formulas with $n = t$.

We next consider the S_k, for $k > 2t$. For certain values of k, this power-sum symmetric function will be a known power of one of the given power-sum symmetric functions:

$$S_k = S_i^{2^j}; \qquad i < k$$

whenever

$$i \cdot 2^j = k \bmod N$$

The values of k which are expressible in this way depend critically on the specific parameters of the code, the block length N and the error-correction capability t. For the perfect Hamming codes ($t = 1$ and N one less than a power of two), there are no odd k which are expressible in the above manner. In general, however, several such k are. For example, for the 5-error-correcting binary BCH code of block length 63, we have S_1, S_3, S_5, S_7 and S_9 given. S_{11}, S_{13} and S_{15} are unknown, but $S_{17} = S_5^{16}$, $S_{19} = S_{13}^{16}$ (which is known); $S_{21} = S_{21}$; S_{23} is unknown, and $S_{25} = S_{11}^{8}$ (which is unknown), $\cdots$, etc.

In general, these relationships may be most readily determined by examining the binary representations of these numbers, as suggested by Mann (Ref. 26). Since multiplication of j by $2 \bmod (2^k - 1)$ is equivalent to a cyclic shift of the k-digit binary expansion of j, S_i is a power of S_j in $GF(2^k)$ if and only if the k-digit binary expansions of i and j are equal except for a cyclic shift.

For every known S_j, j odd, $2t < j < 4t$, we also know $\Delta_{j-2t}^{(t)}$. For each such known S_j, we therefore have an algebraic equation relating the unknown coefficients of the polynomial $f^{(\infty,n)}$. If $\deg \sigma^{(t)} = t$ and there are actually $t + e$ errors, then $\deg f^{(\infty,t)} = e$. Thus, we will have several simultaneous algebraic equations in e unknowns. If we could solve these equations, then we could first determine the polynomial $f^{(\infty,t)}$ and then the error polynomial. Of course, we do not generally know the value of e, but the objective is clearly to solve these equations with a polynomial f of as small degree as possible.

For example, if $\Delta_k^{(t)} = 0$ for all k for which it is known, then the polynomial $f(x) = 1$ solves all of the equations. In this case, of course, the received word lies in a coset containing no more than t errors and the error polynomial is given by $\sigma^{(t)}$.

If instead, $\Delta_k^{(t)} \neq 0$ for some value or values of k, then there must be more than t errors, and $\sigma^{(t)}$ is definitely not the error polynomial. For large values of t, this in itself is well worth knowing, because it enables the decoder to avoid going through a search over the Nth roots of unity to attempt to find the roots of $\sigma^{(t)}$.

In order to correct the additional errors, one must solve the equations for the coefficients of the polynomial f. If there are only $t + 1$ errors, then the algebraic equations contain only one unknown, and the situation is relatively hopeful. If this equation has degree ≤ 4, then it can be quickly solved (without any search) by the methods of Berlekamp, Rumsey and Solomon (SPS 37-39, Vol. IV, pp. 219–226). For example, we saw that in the 5-error correcting code of block length 63, the decoder knows S_1, S_3, S_5, S_7, S_9, from which he can compute $\sigma^{(1)}, \tau^{(1)}, \cdots, \sigma^{(5)}$, $\tau^{(5)}$, $S^{(5)}$ and $A^{(5)}$. Knowledge of $S_{17} = S_5^{16}$ enables one to compute $\Delta_7^{(5)}$, thereby obtaining a quartic equation for $f_1^{(\infty,5)}$. Solution of this quartic enables the decoder to determine $f(x)$, and then the error polynomial $\sigma(x)$, assuming there were no more than 6 errors. If one or more of these equations is linear in some unknown, then that unknown can be eliminated by substitution. For example, if Δ_3 is known, then we can find the error polynomial for a pattern of $(t + 2)$ errors with the solution of a single algebraic equation for f_1. The equation for Δ_3, which is linear in f_2, can be used to eliminate f_2 from the expression for the next known Δ_i, which is algebraic in f_1 and f_2.

In some cases, the equations generated by this method appear to be unnecessarily complicated. For example, consider a two-error-correcting BCH code. The error polynomial can be shown to be given by

$$\sigma(x) = \begin{cases} 1 \text{ if } S_1 = 0 \text{ and } R_3 = S_3 + S_1^3 = 0 & \text{(0 errors)} \\ 1 + S_1 x \text{ if } S_1 \neq 0 \text{ and } R_3 = S_3 + S_1^3 = 0 & \text{(1 error)} \\ 1 + S_1 x + \dfrac{R_3}{S_1} x^2 \text{ if } S_1 \neq 0,\ R_3 \neq 0 \text{ and } Tr\left(\dfrac{R_3}{S_1^3}\right) = 0 & \text{(2 errors)} \\ (1 + (S_1 + \xi) x)\left(1 + \xi x + \left(\dfrac{R_3}{\xi} + S_1^2 + S_1 \xi\right) x^2\right) & \\ \qquad \text{if } R_3 \neq 0 \text{ and } Tr\left(\dfrac{R_3}{S_1^3}\right) \neq 0;\ Tr\left(\dfrac{R_3}{\xi^3}\right) = 0 & \text{(3 errors)} \end{cases}$$

5. Three-Error Case

The zero and one-error cases can be verified immediately. The polynomial for the two-error case follows directly from the iterative algorithm. This polynomial has two roots in the field if and only if Trace $(R_3/S_1^3) = 0$. (For proof, see SPS 37-39, Vol. IV, pp. 219–226.)

In order to verify the three-error case, we must check that the given expression for $\sigma(x)$ satisfies Newton's identities and that it has three distinct roots in the field. By the iterative algorithm a polynomial of fourth degree satisfies Newton's identities when $R_3 \neq 0$ if and only if it is of the form

$$\sigma^{(3)}(x) = 1 + S_1 x + \left(\frac{R_3}{S_1} + \frac{\Delta_1}{R_3}\right) x^2 + \left(\frac{\Delta_1 S_1}{R_3}\right) x^3$$

The assumed polynomial was

$$1 + (S_1 + \xi)\, x \left(1 + \xi x + \left(\frac{R_3}{\xi} + S_1^2 + S_1 \xi\right) x^2\right)$$

These two expressions agree if we set

$$\Delta_1 = \frac{R_3}{S_1}(S_1 + \xi)\left(\frac{R_3}{S} + S_1^2 + S_1 \xi\right)$$

Finally, we must check that the claimed polynomial actually has three roots in the field. The condition for this to occur is that

$$Tr\left(\frac{R_3}{\xi^3}\right) = 0$$

If the minimum weight member of a particular coset is not unique, then there is little gain in decoding that coset, for the *a posteriori* probability of error will be greater than 1/2, no matter how the coset is decoded. For this reason, it is of interest to inquire whether or not the triple error pattern decoded by the above method is unique. The answer depends upon the field. In $GF(2^3)$, there are only three nonzero elements with trace 0; each of them has a unique cube root. Any of the three may be selected as a_0, but, for any given syndrome, the resulting triple error pattern will be the same. The choice of a_0 will affect only those of the three errors denoted by $(S_1 + \xi)$. Since $S_5 = S_1^5$ in $GF(2^3)$, it is evident that this double error correcting BCH code can in fact correct three errors in all cases. Since this code is actually the trivial code consisting only of two codewords, all zeroes and all ones, this is no surprise.

In $GF(2^4)$, there are 7 nonzero elements with trace zero. Three of these are suitable choices for a_1, and three are suitable choices for a_2, but only one of them is a suitable choice for a_0. Thus, in $GF(2^4)$ we must have $a_0 = 1$; there is no other possibility. Hence, if R_3 is a cube, then there are only three possible values of ξ, namely the three cube roots of R_3. The error pattern is represented by $S_1 + \xi_i$, for $i = 1, 2, 3$; with $\xi_i^3 = R_3$. Thus, in $GF(2^4)$, cosets with R_3, a perfect cube, have unique coset leaders. Other cosets of weight three do not.

In all other fields, coset leaders are never unique. In $GF(2^{k+1})$, this is obvious, since there are $2^{2k} - 1$ nonzero elements with trace 0, and any of them may be chosen as a_0; a_1 and a_2 need not be used at all. In $GF(2^{2k})$, we begin by choosing $a_0 = 1$. Since $(2^{2k-1} - 1)$ nonzero field elements have trace zero, and only $(2^{2k-1} - 1)/3$ have multiplicative order divisible by three, it is clear that there are at least $(2^{2k} - 4)/6$ elements with trace 0 and multiplicative order not divisible by three. If one of them provides a suitable choice for a_1, then we may set $a_2 = a_1^2$, or, conversely, if one of them is a suitable choice for a_2, we might set $a_2 = a_1^2$. This shows that there must be an equal number of choices for a_1 and a_2, and hence, at least $(2^{2k-2} - 1)/3$ choices for each.

Finally, we must show that there are multiple choices for the constant a_0. If α is a primitive element of $GF(2^{2k})$, then it can be shown that $a_0 = \alpha^{2(2^k+1)}$ represents such a choice, as does $a_0 = 1$. The proof of this fact is omitted. Thus, there is little value in decoding cosets of weight three in two-error-correcting binary BCH codes, except in the case when the block length is 15.

6. Conclusion

The method introduced here for correcting more than t errors seems to be advantageous chiefly at moderate-to-low information rates (large t, small difference between S_{2t-1} and the next known S_i, and correspondingly small degree of the algebraic equations). The method outlined by Gorenstein, Peterson, and Zierler (Ref. 27) for the special case $t = 2$, and implemented along the lines suggested in (SPS 37-39, Vol. IV, pp. 219–226) has obvious advantages when $t = 2$. At intermediate rates, the method suggested in (Ref. 25) appears preferable.

Our method is seen to be very efficient for correcting one additional error, particularly when S_{2t+3}, S_{2t+5}, or S_{2t+7} is known, since in these cases the algebraic equation for f_1 has degree 2, 3, or 4, respectively. Unfortunately, this method does not appear to be very feasible for correcting more than one additional error, unless someone devises a good algorithm for solving simultaneous algebraic equations in several unknowns in a finite field.

H. Information Processing: Signal and Noise in Nonlinear Devices, *C. A. Greenhall*[12]

This article presents a definition of the signal and noise portions of the output of a rather general nonlinear device. We are able to write down the sample functions of the output signal and noise processes in a particularly simple way. This leads to a formula for the signal output of a hard limiter and thence to a convenient integral expression for the output signal amplitude of a hard bandpass limiter. The input signal can be both amplitude and phase modulated. This integral expression was obtained in the special case of phase modulation by Tausworthe in SPS 37-35, Vol. IV, pp. 307–309. In this article we show the relationship between our method and his.

1. Definition of Output Signal and Noise

Suppose the input to the device in question is

$$x(t) = s(t) + n(t) \qquad (-\infty < t < \infty) \tag{1}$$

where the signal $s(t)$ and noise $n(t)$ are sample functions of independent real stationary processes $S(t, \xi_s)$ and $N(t, \xi_n)$, respectively. The ξ_s and ξ_n are sample "points" (functions) of independent sample spaces Ω_s and Ω_n, which have probability measures P_s and P_n. Assume that these processes have finite variance:

$$E(s^2(t)) = \int_{\Omega_s} S^2(t, \xi_s) P_s(d\xi_s) < \infty$$

$$E(n^2(t)) = \int_{\Omega_n} N^2(t, \xi_n) P_n(d\xi_n) < \infty \tag{2}$$

The output $y(t)\,(-\infty < t < \infty)$ of the device is to be a sample function of a process $Y(t, \xi_s, \xi_n)$ dependent on the input process $x(t)$. We require that the device be time invariant; in other words, if the input is $x(t - \delta)$ then the output is $y(t - \delta)$. This guarantees that Y is stationary. We also ask that

$$E(y^2(t)) = \int_{\Omega_s} \int_{\Omega_n} Y^2(t, \xi_s, \xi_n) P_s(d\xi_s) P_n(d\xi_r) < \infty \tag{3}$$

The fundamental problem is to determine what is meant by the "signal" and "noise" portions of $y(t)$. We write $y(t) = s_y(t) + n_y(t)$, and propose the following conditions on this decomposition:

(i) $s_y(t)$ is to be the sample function of a stationary process $S_y(t, \xi_s)$ of finite variance on the original signal sample space Ω_s.

(ii) $n_y(t) = y(t) - s_y(t)$ is to be uncorrelated with all random variables in the space M of signal random variables of finite variance, i.e., with all random variables $f(\xi_s)$ on Ω_s such that $E(f^2) < \infty$. Thus

$$E(n_y(t) f) = E(n_y(t)) E(f) \tag{4}$$

for all f in M.

Conditions (i) and (ii) imply that

$$E[y(t_1) y(t_2)] = E[s_y(t_1) s_y(t_2)] + E[n_y(t_1) n_y(t_2)] + 2E(n_y) E(s_y) \tag{5}$$

and hence the power spectrum of y is, except possibly for a dc term, the sum of the power spectra of s_y and n_y. This condition is implied in Davenport's paper (Ref. 28) on the bandpass limiter, in which he separates the signal and noise contributions to the power spectrum of the output.

To see how far conditions (i) and (ii) determine s_y and n_y, we write Eq. (4) as

$$E[(n_y - E(n_y)) f] = 0 \tag{6}$$

for all f in M. Then write $y(t)$ as

$$y(t) = [s_y(t) + E(n_y)] + [n_y(t) - E(n_y)]$$

Since $E(n_y)$ is just a constant, it belongs to M. Therefore, $s_y(t) + E(n_y)$ is in M by (i), and $n_y(t) - E(n_y)$ is orthogonal to M by Eq. (6). Thus, $s_y(t) + E(n_y)$ is the projection $p(t) = P(t, \xi_s)$ of the random variable $y(t)$ onto M, the "signal space." An embryonic form of this idea appeared in SPS 37-23, Vol. IV, pp. 160–164, in which the signal portion was defined as the projection

[12] National Research Council postdoctoral resident research associate supported by NASA.

of $y(t)$ onto the subspace generated by just a single random variable in M.

Furthermore, the random variables

$$s_y(t) = p(t) + c$$

$$n_y(t) = y(t) - s_y(t) \qquad (7)$$

where c is any constant, satisfy (i) and (ii). Thus the output signal and noise are determined within constants.

The projection $p(t)$ can be written down explicitly. We know that $p(t)$ is an integrable random variable in Ω_x satisfying $E(p(t)f) = E(y(t)f)$ for all bounded measurable f on Ω_x. This implies that with probability one on Ω_x,

$$P(t, \xi_x) = E(y(t)|S)(\xi_x) = \int_{\Omega_n} Y(t, \xi_x, \xi_n) P_n(d\xi_n) \qquad (8)$$

the conditional expectation of $y(t)$ with respect to the original signal process $S(t, \xi_x)$. It can be verified that P is stationary.

We now set $c = 0$ in Eq. (7) and adopt our definitions of output signal and noise

$$s_y(t) = E(y(t)|S)$$

$$n_y(t) = y(t) - s_y(t) \qquad (9)$$

In summary, the signal portion of the output at time t is obtained by fixing the input signal and averaging the output at time t over all possible noise inputs belonging to the noise process N.

The definition Eq. (9) may also be of use for nonstationary input signals. If the processes are not stationary, then conditions (i) and (ii) yield that

$$s_y(t) = E(y(t)|S) + c(t)$$

where $c(t)$ is an arbitrary deterministic function of time. Some further condition (like c = constant) is needed to define s_y well enough.

2. Special Classes of Nonlinear Devices

We apply the definitions Eq. (9) of output signal and noise to the following two classes of devices.

a. Linear filter. Let

$$y(t) = (Hx)(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)\,d\tau \qquad (-\infty < t < \infty) \qquad (10)$$

where $x = s + n$ as in Eq. (1) and

$$\int_{-\infty}^{\infty} (1 + t^2) h^2(t)\,dt < \infty \qquad (11)$$

The condition Eq. (11) on the impulse response h of the filter ensures that with probability one the integral Eq. (10) exists for all t and that $E(y^2) < \infty$, given that $E(x^2) < \infty$. We will show that

$$s_y(t) = (Hs)(t), \qquad n_y(t) = (Hn)(t) \qquad (12)$$

where s_y and n_y are defined by Eq. (9). The condition Eq. (11) and $E(s^2) < \infty$ ensures that $(Hs)(t)$ is a random variable on Ω_s with finite variance, so all we have to do is verify the projection property

$$E[(Hs)(t)f] = E[y(t)f] \qquad (13)$$

for all random variables f on Ω_s of finite variance. Thus

$$E\left[f\int_{-\infty}^{\infty} h(t-\tau)x(\tau)\,d\tau\right] = \int_{-\infty}^{\infty} h(t-\tau)E[f(s(t)+n(t))]\,d\tau$$

$$= \int_{-\infty}^{\infty} h(t-\tau)E[fs(t)]\,d\tau = E\left[f\int_{-\infty}^{\infty} h(t-\tau)s(\tau)\,d\tau\right] \qquad (14)$$

which is Eq. (13). With the definition Eq. (9), then, a linear filter does not mix the signal and noise.

This property extends further. Suppose we follow the nonlinear device of subsection 1 by a filter H. Thus the combined output is $z(t) = (Hy)(t) = (Hs_y)(t) + (Hn_y)(t)$. If we replace s and n in Eq. (14) by s_y and n_y, and note that $E(fn_y(t)) = 0$ by definition, we see that

$$s_{Hy}(t) = Hs_y(t), \qquad n_{Hy}(t) = Hn_y(t) \tag{15}$$

The decomposition Eq. (9) is not affected by passing through a linear filter.

b. Zero-memory device. Here we let

$$y(t) = F(x(t)) \tag{16}$$

where F, the characteristic of the device, is a real-valued function such that

$$E[y^2(t)] = \int_{\Omega_s}\int_{\Omega_n} F^2(S(t,\xi_s) + N(t,\xi_n))\,P_s(d\xi_s)\,P_n(d\xi_n) < \infty$$

According to Eqs. (8) and (9),

$$\begin{aligned} s_y(t) &= \int_{\Omega_n} F(s(t) + N(t,\xi_n))\,P_n(d\xi_n) \\ &= \int_{-\infty}^{\infty} F(s(t) + n)\,p(n)\,dn = G(s(t)) \end{aligned} \tag{17}$$

where p is the probability density of $n(t)$. As far as the signal is concerned, the device acts like another device with characteristic G, which, of course, depends on the noise distribution $p(n)\,dn$. We will call G the ***signal characteristic*** of the device.

Now suppose the input signal and noise to this device F are narrow-band about a center frequency ω_0. Let the signal have the form $s(t) = V(t)\sin(\omega_0 t + \theta(t)), V(t) \geqq 0$, where $V\sin\theta$ and $V\cos\theta$ are narrow-band about zero frequency with bandwidths small compared with ω_0. Also, assume that the random variables $V(t)$ and $\theta(t)$, t fixed, are independent, that $\theta(t)$ is uniformly distributed in $[0, 2\pi]$, and that the distribution of $V(t)$ is independent of t. Then

$$\begin{aligned} E[G^2(s(t))] &= E[G^2(V(t)\sin(\omega_0 t + \theta(t)))] \\ &= E\left[\frac{1}{2\pi}\int_0^{2\pi} G^2(V(t)\sin\phi)\,d\phi\right] < \infty \end{aligned} \tag{18}$$

Hence with probability one, we can expand $G(V(t)\sin\phi)$ in a Fourier series:

$$G(V(t)\sin\phi) = \sum_{k=-\infty}^{\infty} c_k(V(t))\,e^{ik\phi} \tag{19}$$

in $L^2(0, 2\pi)$, where the Fourier coefficients c_k are given by

$$c_k(V) = \frac{1}{2\pi}\int_0^{2\pi} G(V\sin\phi)\,e^{-ik\phi}\,d\phi \tag{20}$$

The same change of variable as was used in Eq. (18) will give that

$$G(s(t)) = \sum_{k=-\infty}^{\infty} c_k(V(t))\,e^{ik(\omega_0 t + \theta(t))} \tag{21}$$

in $L^2(\Omega_s)$. The convergence in Eq. (21) is uniform in t because the distribution of $V(t)$ is independent of t, by assumption. The terms in Eq. (21) are the signal components in the narrow frequency zones about each $\pm k\omega_0$ $(k = 1, 2, \cdots)$. We can find a (nonrealizable) filter H satisfying Eq. (11) whose complex transfer function is 1 in the kth zone and 0 in all other zones (by making the transfer function smooth enough). If such a filter passes the kth harmonic unchanged and annihilates all the others, then by Eq. (15) the signal output of the device consisting of the zero-memory device followed by H is

$$2\,Re\,[c_k(V(t))\,e^{ik(\omega_0 t + \theta(t))}] \tag{22}$$

In the first zone, the phase modulation $\theta(t)$ passes through unchanged.

3. Bandpass Limiter

An example of a zero-memory device is the ideal bandpass limiter, where

$$F(x) = 1 \qquad (x \geq 0)$$

$$= -1 \qquad (x < 0) \qquad (23)$$

(See Eq. 16.) Henceforth the input noise $n(t)$ will be a stationary Gaussian process, with $E(n) = 0$, $E(n^2) = \sigma^2$. We easily calculate from Eq. (17) that the signal at the output of the hard limiter is

$$s_{\prime\prime}(t) = G(s(t)) = g\left(\frac{s(t)}{\sigma}\right)$$

$$g(x) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^x e^{-\frac{1}{2}z^2} dz \qquad (24)$$

(See Fig. 23.) For large signal-to-noise ratios the signal characteristic G is itself like a hard limiter. For small signal-to-noise ratios, G is almost linear, i.e.,

$$G(s) \approx \left(\frac{2}{\pi}\right)^{1/2} \frac{s}{\sigma} \qquad (25)$$

Consider now the case of narrow-band signal and noise inputs. The harmonic expansion Eq. (21) of $s_y(t)$ can be written

$$s_y(t) = \sum_{k=1}^{\infty} b_k(v(t)) \sin(k\omega_0 t + k\theta(t)), \; v(t) = \frac{V(t)}{\sigma} \qquad (26)$$

$$b_k(v) = \frac{1}{\pi} \int_{-\pi}^{\pi} g(v \sin\phi) \sin k_\psi \, d\phi \qquad (k = 1, 2, \cdots) \qquad (27)$$

if k is even then $b_k = 0$.

Tausworthe (SPS 37-35, Vol. IV, pp. 307–309) obtained the expression Eq. (27) for the signal amplitude in the kth zone, in the absence of amplitude modulation. It can be put in the form of a hypergeometric function, Davenport's original form (Ref. 28). Here is a Bessel function form of b_k: integrate Eq. (27) by parts to obtain

$$\begin{aligned}
b_k(v) &= \left(\frac{2}{\pi}\right)^{1/2} \frac{v}{k\pi} \int_{-\pi}^{\pi} e^{-\frac{1}{2}v^2 \sin^2\phi} \cos k\phi \cos\phi \, d\phi \\
&= \left(\frac{2}{\pi}\right)^{1/2} \frac{v}{k\pi} \int_{-\pi}^{\pi} e^{-\frac{1}{4}v^2(1-\cos 2\phi)} \frac{1}{2} [\cos(k-1)\phi + \cos(k+1)\phi] \, d\phi \\
&= \left(\frac{2}{\pi}\right)^{1/2} \frac{v}{k} e^{-\frac{1}{4}v^2} \left[\frac{1}{4\pi} \int_{-2\pi}^{2\pi} e^{\frac{1}{4}v^2 \cos\theta} \cos\left(\frac{k-1}{2}\theta\right) d\theta + \frac{1}{4\pi} \int_{-2\pi}^{2\pi} e^{\frac{1}{4}v^2 \cos\theta} \cos\left(\frac{k+1}{2}\theta\right) d\theta\right] \\
&= \left(\frac{2}{\pi}\right)^{1/2} \frac{v}{k} e^{-\frac{1}{4}v^2} \left[I_{\frac{1}{2}(k-1)}\left(\frac{1}{4}v^2\right) + I_{\frac{1}{2}(k+1)}\left(\frac{1}{4}v^2\right)\right] \qquad (k \text{ odd}) \qquad (28)
\end{aligned}$$

where the I_m are modified Bessel functions of the first kind.

The signal power in the kth zone is

$$\frac{1}{2} E\left[b_k^2\left(\frac{V(t)}{\sigma}\right)\right]$$

From this we can compute signal-to-noise ratio in this zone, since the total power there is $8/(\pi k)^2$ (SPS 37-35, Vol. IV, pp. 307–309), and the signal and noise are uncorrelated.

4. Comparison with Tausworthe's Treatment

We will briefly review Tausworthe's 1965 (SPS 37-35, Vol. IV, pp. 307–309) derivation of the signal amplitude Eq. (27) in the kth zone of the output of a hard limiter. He assumes a signal

$$s(t) = (2)^{1/2} \sin(\omega_0 t + \theta(t)) \qquad (29)$$

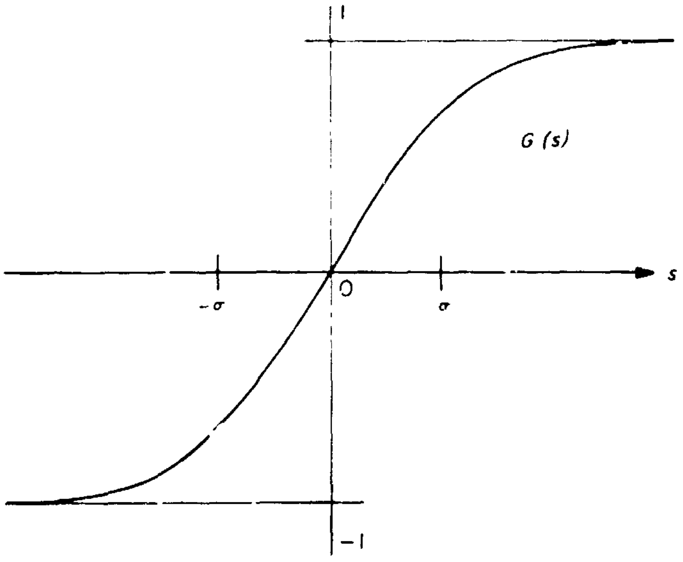

Fig. 23. Signal characteristic of a hard limiter in Gaussian noise

where $\theta(t)$ is uniformly distributed. The noise is narrow band stationary Gaussian. The output of the hard limiter is written

$$y(t) = \sum_{k=1,\,k \text{ odd}}^{\infty} y_k(t)$$

$$y_k(t) = S_k(t) + N_k(t), S_k(t) = E(y(t)\,w_k(t))\,w_k(t)$$

$$w_k(t) = (2)^{1/2} \sin(k\,\omega_0 t + k\theta(t)) \qquad (30)$$

where y_k, S_k, N_k are, respectively, the total output, signal, and noise in the kth zone. He then puts $E(y(t)\,w_k(t))$ into the form of Eq. (27).

To see why our methods agree, first we remark that

$$E(y_k(t)\,w_l(t)) = 0 \qquad (k \neq l; k, l = 1, 2, \cdots) \qquad (31)$$

This can be shown by writing $y_k(t)$ in the form $A(t)\sin(k\omega_0 t + k\psi(t))$ (SPS 37-35, Vol. IV, pp. 307–309, section 3), and observing that the distribution of $\psi(t)$ given $\theta(t)$ is symmetric about $\theta(t)$. The calculations are omitted.

Summing over all zones, the output signal is

$$S(t) = \sum_{k=1}^{\infty} E(y(t)\,w_k(t))\,w_k(t) \qquad (32)$$

where we have used Eq. (31) to justify including the even harmonics. Thus $S(t)$, t fixed, is just the projection of the random variable $y(t)$ onto the subspace generated by $w_1(t), w_2(t), \cdots$. By the same maneuver as was used in Eq. (16) it can be shown that this subspace is just the set of random variables $f(s(t))$ depending only on $s(t)$ such that $E[f^2(s(t))] < \infty$. However, our $s_y(t)$ in Eq. (26) also belongs to this subspace and is the projection of $y(t)$ onto the larger subspace of random variables which depend on the whole signal process. Therefore, $S(t) = s_y(t)$; i.e., our methods give the same signal component of the output when $V(t)$ is constant.

I. Data Compression Techniques: Use of Six and Eight Quantiles to Test Hypotheses in Data-Compressed Experiments, *I. Eisenberger*

1. Introduction

Theoretical analyses relative to the use of sample quantiles for data compression of space telemetry have been given in three previous JPL Technical Reports, Refs. 29–31. The first of these reports, Ref. 29, deals with the problem of estimating the parameters of a normal distribution using up to twenty sample quantiles, and also describes two goodness-of-fit tests, each using four sample quantiles. The second and third reports, Refs. 30 and 31 are concerned with hypothesis testing and the estimation of the correlation coefficient ρ of a bivariate normal distribution, using up to four sample quantiles. A forthcoming technical report, *Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles, III*, will extend most of the results given in Refs. 30 and 31 to six and eight sample quantiles. The purpose of this article is to give the hypotheses and assumptions relating to the tests and the assumptions relating to estimating ρ. The derivation and statement of the test statistics and estimators are given in the report.

2. Review of Quantiles

To define a quantile, consider a sample of n independent values $x_1, x_2, \cdots, x_n$ taken from a distribution of continuous type with distribution function $G(x)$ and density function $g(x)$. *The pth quantile*, or the *quantile of order p* of the distribution or population, denoted by ζ_p, is defined as the root of the equation $G(\zeta) = p$; that is

$$p = \int_{-\infty}^{\zeta_p} dG(x) = \int_{-\infty}^{\zeta_p} g(x)\,dx$$

The corresponding *sample* quantile z_p is defined as follows: If the sample values are arranged in nondecreasing order of magnitude

$$x_{(1)} \leqq x_{(2)} \leqq \cdots \leqq x_{(n)}$$

then $x_{(i)}$ is called the ith order statistic and

$$z_p = x_{[np]+1}$$

where $[np]$ is the greatest integer $\leqq np$.

If $g(x)$ is differentiable in some neighborhood of each quantile value considered, it has been shown (Ref. 32) that the joint distribution of any number of sample quantiles is asymptotically normal as $n \to \infty$ and that, asymptotically,

$$E(z_p) = \zeta_p$$

$$\mathrm{Var}(z_p) = \frac{p(1-p)}{n\,g^2(\zeta_p)}$$

$$\rho_{12} = \left[\frac{p_1(1-p_2)}{p_2(1-p_1)}\right]^{1/2}$$

where ρ_{12} is the correlation between z_{p_1} and z_{p_2}, $p_1 < p_2$. The statement "$g(x) = N(\mu, \sigma)$" will mean that the random variable under consideration is normally distributed with mean μ, variance σ^2, and has the density function $g(x)$ associated with it.

3. Hypotheses and Assumptions

For comparison purposes, the designation here of the tests will coincide with that used for the tests in the reports.

In test A, we are given a set of n independent observations from a normal population with known variance σ^2; the test is designed to decide whether the mean μ has a value of μ_1 or a value of μ_2. In test $\bar{\text{A}}$, the assumption that σ^2 is known is not used.

In test B, we test whether the standard deviation σ has a value of σ_1 or a value of σ_2. When more than one quantile is used, it is not necessary to assume that μ is known.

In tests D, $\bar{\text{D}}$, and $\bar{\text{E}}$, it is assumed that we are given sets of independent samples taken from two independent normal populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. In test D we assume that $\sigma = \sigma_1 = \sigma_2$ is known and μ_1 is unknown; we test whether $\mu_2 = \mu_1$ or $\mu_2 = \mu_1 + \theta$, $\theta \neq 0$. In test $\bar{\text{D}}$, the assumption that σ is known is not used. In test $\bar{\text{E}}$ we assume that both $\mu = \mu_1 = \mu_2$ and σ_1 are unknown and test whether $\sigma_2 = \sigma_1$ or $\sigma_2 = \theta\sigma_1$, $\theta > 0$.

In tests F and $\bar{\text{F}}$, we are given n independent pairs of observations $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ taken from two normally distributed populations where $g_1(x) = N(\mu_1, \sigma_1)$ and $g_2(y) = N(\mu_2, \sigma_2)$. In test F we assume that μ_1, μ_2, σ_1, σ_2 are known and test whether $\rho = 0$ or $\rho \neq 0$. In test $\bar{\text{F}}$ we assume that both $\mu = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$ are unknown and again test whether $\rho = 0$ or $\rho \neq 0$.

In estimating ρ, it will first be assumed that the assumptions of test F hold. This estimator will be denoted by $\hat{\rho}_1$. For the second estimator $\hat{\rho}_2$ it will be assumed that $\mu = \mu_1 = \mu_2$ is unknown and that σ_1 and σ_2 are known.

Table 3 summarizes the above hypotheses and assumptions.

Table 3. Hypotheses and assumptions relating to the tests, and assumptions relating to estimating ρ_1 and ρ_2

Test	Null hypothesis	Alternative hypothesis	Assumptions
A	$g(x) = N(\mu_1, \sigma)$	$g(x) = N(\mu_2, \sigma)$	σ known
$\bar{\text{A}}$			σ unknown
B	$g(x) = N(\mu, \sigma_1)$	$g(x) = N(\mu, \sigma_2)$	μ unknown
D	$g_1(x) = N(\mu, \sigma)$	$g_1(x) = N(\mu, \sigma)$	x and y independent; σ known, μ unknown.
$\bar{\text{D}}$	$g_2(y) = N(\mu, \sigma)$	$g_2(y) = N(\mu + \theta, \sigma)$ $\theta \neq 0$	x and y independent; μ and σ unknown.
$\bar{\text{E}}$	$g_1(x) = N(\mu, \sigma)$ $g_2(y) = N(\mu, \sigma)$	$g_1(x) = N(\mu, \sigma)$ $g_2(y) = N(\mu, \theta\sigma)$ $\theta > 0$	x and y independent; μ and σ unknown.
F	$g_1(x) = N(\mu_1, \sigma_1)$	$g_1(x) = N(\mu_1, \sigma_1)$	$\mu_1, \mu_2, \sigma_1, \sigma_2$ known
$\bar{\text{F}}$	$g_2(y) = N(\mu_2, \sigma_2)$ $\rho = 0$	$g_2(y) = N(\mu_2, \sigma_2)$ $\rho \neq 0$	$\mu_1 = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$ unknown
	Estimating ρ_1		$\mu_1, \mu_2, \sigma_1, \sigma_2,$ known
	Estimating ρ_2		$\mu_1 = \mu_2 = \mu$ unknown σ_1 and σ_2 known

J. Astrometrics: A New Method For Extracting The Reflectivity Distribution From Planetary Radar Data, *S. Zohar*

1. Introduction

The present approach to the extraction of the planetary reflectivity distribution from the range gated reflection off the observed planet is based on the spectrum of this signal. In practice, the spectrum is computed from the signal's measured autocorrelation. A recently derived result (SPS 37-43, Vol. IV, pp. 330–338) makes it possible to extract the information directly from the autocorrelation function. Computationally, this method is markedly different from the spectral approach, being based on a different set of assumptions. There are indications that the new method will more closely approximate reality. However, this is difficult to substantiate on theoretical grounds; only test case comparisons between the two methods will tell which is better. In the meantime, the main importance of this new method lies simply in its providing a different path to the desired reflectivity distribution. Thus, any feature or details which show up in both processing schemes could be considered with a high degree of confidence as representing physical reality.

2. Formulation of the Problem

Let the reflectivity distribution be subjected to a parallel projection along the line of sight onto a plane perpendicular to this line. We introduce the coordinates θ, ν in the projection plane as shown in Fig. 24. These are associated with range and doppler shift, respectively. The reflectivity distribution can now be expressed as a two dimensional Fourier series in these two variables.[13]

$$\hat{F}(\theta, \nu) = \sum_{n, r = -\infty}^{\infty} b_{nr} e^{i2(\pi n \nu + r\theta)}$$

with

$$b_{-n,r} = b_{nr}; \qquad b_{n,-r} = b_{nr}. \tag{1}$$

We have shown (SPS 37-43, Vol. IV, pp. 330–338) that the Nyquist rate samples of the (normalized) range-gated autocorrelation function of the reflected signal (ρ_k) depend on the b_{nr} Fourier coefficients as follows:

$$\rho_k = h_0 \frac{W_T}{W} \sum_{k, n, r = -\infty}^{\infty} a_k b_{nr} \cos [2\pi (n + k) \nu_c] \times J_{k+r}^2 [\pi (n + k) \hat{\nu}]; \tag{2}$$

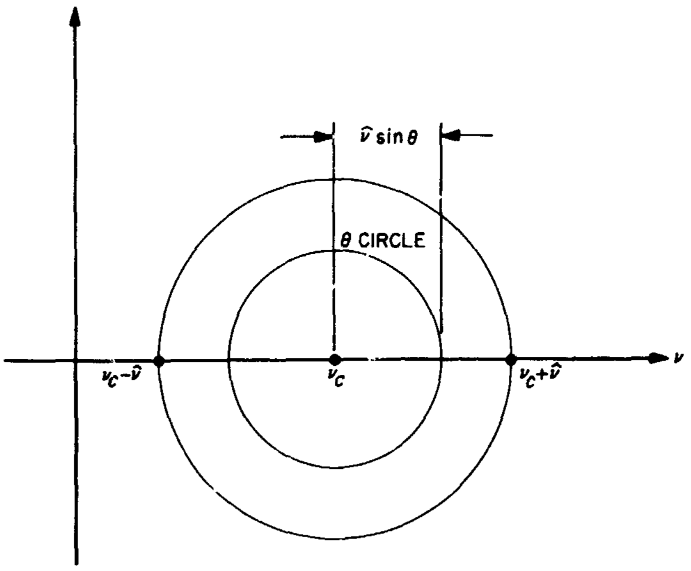

Fig. 24. Coordinates for the reflectivity distribution

$\hat{\nu}$ appearing in the Bessel function argument is half the normalized limb-to-limb bandwidth; ν_c is the normalized frequency of the subradar point (Fig. 24). Here h_0, a_k are computable from the range-gate and backscatter functions, while W_T and W are, respectively, computable and measurable powers in the radar system.

We have also shown that for a featureless "grey planet" Eq. (2) leads to the much simpler samples g_k given by:[14]

$$g_k = \cos (2\pi k \nu_c) \sum_{k=0}^{\infty} \varepsilon_k a_k J_k^2 (\pi k \hat{\nu}). \tag{3}$$

Eqs. (2) and (3) are the basis for the new extraction procedure.

Our purpose is to find the reflectivity distribution $\hat{F}(\theta, \nu)$, given the correlation coefficients ρ_k. Equivalently, our problem can be referred to as the inversion of Eq. (2); namely, given the set ρ_k, find the set b_{nr}. It should be realized from the outset that if only the first I samples are available, then there is no unique $\hat{F}(\theta, \nu)$ that fits them; rather, there is a large family of possible solutions, all of which share the first I correlation coefficients. We can narrow the range of possible functions to a single one only by introducing specific assumptions to compensate for the missing data. What assumptions should one make? Here we have only some simple guidelines, such as the requirement that the assumptions should be plausible and that the solutions based on different assumptions should not be

[13] Strictly speaking, $\hat{F}(\theta, \nu)$ is not the reflectivity function defined in SPS 37-43, Vol. IV, pp. 330–338; it is the so-called "feature function." The distinction between the two, however, is not important here; hence, we use somewhat loosely the term "reflectivity distribution." For details, as well as the justification of Eq. (1), see SPS 37-43, Vol. IV, pp. 330–338.

[14] ε_k is Neumann's symbol ($\varepsilon_k = 2$, except $\varepsilon_0 = 1$).

drastically different. In the final analysis, of course, the assumptions should lead to manageable mathematics.

It is with this background in mind that we approach the inversion of Eq. (2). Our fundamental assumption is that all b_{nr}'s vanish for n, r greater than some specific bounds consistent with the amount of data available. The basic justification for this is that $\hat{F}(\theta, \nu)$, being a function representing some aspect of physical reality, is not expected to have jumps or discontinuities. Its Fourier coefficients b_{nr} should, therefore, decay to zero with increasing n, r.

The effect of this assumption on Eq. (2) is to replace the infinite summations over n, r with finite ones so that our problem is essentially reduced to the solution of a finite set of linear equations. A possible approach to the problem is to obtain the autocorrelation for several adjacent zones covering the region of interest and then process the data of all the zones simultaneously. Such a procedure does actually lead to a reasonable estimate of $\hat{F}(\theta, \nu)$, and we shall return to it later on. However, the amount of computation involved in such a scheme is prohibitive and seems to rule this approach out as a practical solution.

An alternative approach which does lead to a practical scheme is to use sufficiently narrow gates so that $\hat{F}(\theta, \nu)$ could be assumed independent of θ over the zone. This results in the processing of each zone independently of the others, leading to a drastic reduction in computations. We consider this approach in detail in the next subsection.

3. The New Extraction Method

Our initial task is to obtain the special case of Eq. (2) for the constraint that $\hat{F}(\theta, \nu)$ be independent of θ. This constraint is satisfied by assigning the following form to b_{nr}:

$$b_{nr} = b_{n0}\delta_{r0} \equiv b_n \delta_{r0} .$$

Substitution of this in Eq. (2) leads to[15]

$$\rho_k = h_0 \frac{W_T}{W} \sum_{n=-\infty}^{\infty} b_n \cos\left[2\pi (k+n) \nu_c\right]$$

$$\times \sum_{k=0}^{\infty} \varepsilon_k a_k J_k^2 \left[\pi (k+n) \hat{\nu}\right\} .$$

This, however, can be expressed in terms of the correlation coefficients of the corresponding "grey planet" Eq. (3), yielding

$$\rho_k = h_0 \frac{W_T}{W} \sum_{n=-\infty}^{\infty} g_{k-n} b_n . \qquad (4)$$

The computation of the g-parameters has been discussed in detail (SPS 37-43, Vol. IV, pp. 330–338). Of the remaining parameters, W_T and h_0 can be computed; whereas W, the signal power, can be measured. We see, therefore, that if the ρ_k's are measured, the only unknowns in Eq. (4) are the b_n's. In practice, however, we never get the autocorrelation of the received signal directly. The receiver, from whose output we try to obtain the ρ_k's, will usually have a nonflat frequency response, as well as its own noise in the frequency band of interest. Under these conditions, the measured autocorrelation is not the signal autocorrelation appearing in Eq. (4). It is, however, simply related to it. Thus, it has been shown (SPS 37-31, Vol. IV, pp. 315–320) that ρ_k is related to measured parameters by the following equation:

$$\beta m_i - n_i = \alpha \sum_{k=-\infty}^{\infty} n_{i-k} \rho_k \qquad (5)$$

where

n_i = ith normalized correlation coefficient of the receiver output when fed with white noise;

m_i = ith normalized correlation coefficient of the receiver output when a signal is received;

$$\alpha = \frac{W}{W'_N} = \frac{\text{signal power at receiver input}}{\text{noise power at receiver input}} ;$$

$$\beta = \frac{W_M}{W_N} = \frac{\text{signal plus noise power at receiver output}}{\text{noise power at receiver output}} .$$

Obviously, the equation relating the unknown b_n's to measured entities is obtained by substituting Eq. (4) in Eq. (5). This yields

$$v_i \equiv \frac{1}{h_0} \cdot \frac{W_{N'}}{W_T} (\beta m_i - n_i) = \sum_{k,j=-\infty}^{\infty} n_{i-k} g_{k-j} b_j . \qquad (6)$$

We pause here briefly to point out an interesting aspect of Eq. (6). The basic fact leading Eq. (5) is that for band-

[15] The vanishing of the high order b_n's will be handled later on.

limited signals, the effect of a filter on the autocorrelation can always be represented in the form

$$y_i = \sum_{k=-\infty} p_{i-k} x_k$$

where

x_k = autocorrelation coefficient of input signal;

y_i = autocorrelation coefficient of output signal;

p_j = a set of parameters characterizing the filter and derivable from its more familiar transfer function.

In view of this, we may rephrase Eq. (6)

$$v_i = \sum_{k=-\infty}^{\infty} n_{i-k} \left(\sum_{j=-\infty}^{\infty} g_{k-j} b_j \right)$$

and take the attitude that the v_i's are the autocorrelation coefficients of a signal obtained by passing the "feature signal" (the "signal" whose correlation coefficients are the b_j's) through two filters in cascade: the "spherical filter" (parameters g_i) and the filter representing the receiver (parameters n_i) (Fig. 25).

From this point of view it is quite clear that obtaining the feature function is essentially identical with the more familiar problem of obtaining a receiver input autocorrelation from its measured output autocorrelation. The only difference is that we have here a "receiver" with a more complicated (but computable) transfer function.

We return now to the solution of Eq. (6). Invoking our assumption concerning the vanishing of the high order b_j's, we set

$$b_j = 0 \qquad \text{for} |j| > J. \tag{7}$$

Hence

$$v_i = \sum_{j=-J}^{J} \left\{ \sum_{k=-\infty}^{\infty} n_{i-k} g_{k-j} \right\} b_j = \sum_{j=-J}^{J} S_{ij} b_j \tag{8}$$

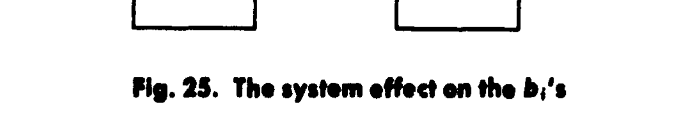

Fig. 25. The system effect on the b_i's

where

$$S_{ij} = \sum_{k=-\infty}^{\infty} n_{i-k} g_{k-j}. \tag{9}$$

From Eq. (8) it is obvious that to solve for the b_j's we must have all v_i for $|i| \leq J$. The solution would then involve the inversion of the matrix S (whose ij element is S_{ij}) of order $(2J + 1)$. We turn our attention now to this matrix.

Usually, only a finite number of n_k's is available. Hence, as before, we compensate for our incomplete information by setting the unavailable n_k's to zero. Conceptually, however, the situation here is different from the previous one. The transfer function of the receiver could, in principle, be approximated quite well by a rational frequency function. From the finite set of parameters of this rational function, one could compute the infinite set of n_k's. This would then require the computation of an infinite number of g_i's. Thus, the motivation for the truncation of the k summation in Eq. (9) is not so much a lack of data as it is a desire to obtain a practical computation scheme.

The truncation is effected by setting

$$n_k = 0 \qquad \text{for} |k| > K.^{16}$$

Hence

$$S_{ij} = \sum_{k=i-K}^{i+K} n_{i-k} g_{k-j}. \tag{10}$$

We proceed now to prove that S is a symmetric Toeplitz matrix.[17] This is simply effected by setting $k - j = q$ in Eq. (10). Thus

$$S_{ij} = \sum_{q=(i-j)-K}^{(i-j)+K} n_{(i-j)-q} g_q = S_{i-j}. \tag{11}$$

We see that the dependence of S_{ij} on i, j is only through the difference $(i - j)$. Hence, the matrix S is a Toeplitz matrix.

[16] Present JPL practice is to use the same equipment to obtain both m_i and n_i. Hence $K = J$.

[17] A matrix μ whose ij element μ_{ij} is a function of $i - j$ rather than i, j separately will have one and the same element along every diagonal, superdiagonal or subdiagonal. It is known as a Toeplitz matrix.

To prove symmetry we rewrite Eq. (11) as follows

$$S_k = \sum_{q=k-K}^{k+K} n_{k-q} g_q .$$

Hence

$$S_{-k} = \sum_{q=-(k+K)}^{K-k} n_{-k-q} g_q .$$

The evenness of the autocorrelation functions $n(\tau), g(\tau)$ now yields

$$S_{-k} = \sum_{j=k-K}^{k+K} n_{k-j} g_j = S_k .$$

We conclude that to solve Eq. (8) and obtain the b_j's we have to invert a symmetric Toeplitz matrix of order $(2J + 1)$. This is readily accomplished even for large J by the method of W. F. Trench (Ref. 33). Preliminary analysis indicates, for example, that the complete solution of Eq. (3) for $J = 50$ would take about 4 sec on the IBM 7094.

We turn now briefly to the general case mentioned earlier in which we do not neglect the θ variation of the feature function across the observed zone. Instead of having to deal with one autocorrelation function, we now consider $(Z + 1)$ such functions corresponding to $(Z + 1)$ different zones. To distinguish among these we attach an extra index to ρ, h, a, w. Thus, we have for zone z

$$\rho_{kz}, h_{0,z}, a_{kz}, W_z .$$

With this terminology, Eq. (2) becomes

$$\rho_{kz} = h_{0z} \frac{W_T}{W_z} \sum_{k,n,r=-\infty}^{\infty} b_{nr} a_{kz} \cos [2\pi (n + k) \nu_c] J^2_{k+r} [\pi (n + k)\hat{\nu}]$$

$$(z = 0, 1, \cdots, Z) . \qquad (12)$$

In analogy with the g_m-parameters introduced in the special case, we introduce now the Y_{mrz}-parameters defined as follows:

$$Y_{mrz} = h_{0z} \frac{W_T}{W_z} \cos (2\pi m \nu_c) \sum_{k=-\infty}^{\infty} a_{kz} J^2_{k+r} (\pi m \hat{\nu})$$

$$= \quad \frac{1}{2} h_{0z} \frac{W_T}{W_z} \cos (2\pi m \nu_c) \sum_{k=0}^{\infty} \varepsilon_k a_{kz} [J^2_{r+k} (\pi m \hat{\nu}) + J^2_{r-k} (\pi m \hat{\nu})] .$$

Applying this to Eq. (12) we get

$$\rho_{kz} = \sum_{n,r=-\infty}^{\infty} Y_{k-n,r,z} b_{nr} .$$

We note that Y_{mrz} is an even function of r. Combining this with the symmetry properties of b_{nr} (SPS 37-43, Vol. IV, pp. 330–338), we get

$$\rho_{kz} = \frac{1}{2} \sum_{n,r=0}^{\infty} \varepsilon_n \varepsilon_r (Y_{k-n,r,z} + Y_{k+n,r,z}) b_{nr} .$$

Finally, in analogy with Eq. (7) we assume

$$b_{nr} = 0 \qquad \text{for } |n| > I \text{ or } |r| > Z ,$$

where I is the number of available correlation coefficients in each of the $Z + 1$ zones. This leads to

$$\rho_{kz} = \frac{1}{2} \sum_{n,r=0}^{n=I,\; r=Z} \epsilon_n \epsilon_r (Y_{k-n,r,z} + Y_{k+n,r,z})\, b_{nr}. \tag{13}$$

This equation can be solved by the simple artifice of arranging b_{nr} and ρ_{k_z} as column matrices. Specifically, let us introduce the following two column matrices of order $I + 1$

$$\hat{\rho}_z = \begin{bmatrix} \rho_{0,z} \\ \rho_{1,z} \\ \cdot \\ \cdot \\ \cdot \\ \rho_{I,z} \end{bmatrix} \quad ; \quad \hat{b}_r = \epsilon_r \begin{bmatrix} b_{0,r} \\ b_{1,r} \\ \cdot \\ \cdot \\ \cdot \\ b_{I,r} \end{bmatrix}$$

and the square matrix $\hat{Y}_{rz}$ of order $(I + 1)$ whose k, n element is given by

$$(\hat{Y}_{rz})_{kn} = \frac{\epsilon_n}{2} (Y_{k-n,r,z} + Y_{k+n,r,z}).$$

Eq. (13) can now be expressed in terms of these matrices as follows:

$$\begin{bmatrix} \hat{\rho}_0 \\ \hat{\rho}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\rho}_Z \end{bmatrix} = \begin{bmatrix} \hat{Y}_{00} & \hat{Y}_{01} & \cdots & \hat{Y}_{0Z} \\ \hat{Y}_{10} & \hat{Y}_{11} & \cdots & \hat{Y}_{1Z} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \hat{Y}_{Z0} & \hat{Y}_{Z1} & \cdots & \hat{Y}_{ZZ} \end{bmatrix} \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{b}_Z \end{bmatrix}$$

We see that solving for the $(I + 1)(Z + 1)$ required b_{nr} coefficients calls for the inversion of a matrix of order $(I + 1)(Z + 1)$.

No special exploitable structure is evident here and the matrix is not even symmetric. It should be pointed out in passing that there is an alternative formulation in which each submatrix turns out to be a Toeplitz matrix though of almost twice the order $(2I + 1)$. However, it is not obvious how such a structure could be exploited to simplify the inversion.

In the present JPL system, a typical case would involve $I = 50$ $Z = 8$. Hence, the matrix to be inverted is of order 459. It is doubtful whether such an inversion should even be attempted in view of the accumulation of errors and prohibitive cost.

Attacking the same problem with the previous scheme, we would treat each of the 9 zones separately (neglecting the θ variation inside a zone) and end up by inverting 9 symmetric Toeplitz matrices of order 101. Solving all 9 matrix equations would take about 33 sec on the IBM 7094.

K. Astrometrics: Elimination of Charged Electron Effects from Ranging Data, *P. E. Reichley*

1. Summary

The effect of charged electrons on ranging data in the ionosphere and interplanetary medium plays an important role in the growing need for greater tracking accuracies. We present in this article a method of obtaining the range change from tracking point to tracking point with effects due only to the troposphere. The effects due to the charged electron distribution appear as second order effects in this method. These results are obtained under the assumption that the troposphere and charged electron distributions are spherically symmetric and that the spacecraft is being tracked with both doppler and range points (obtained by a ranging code). This range change can then be utilized to yield the true range with or without a constant bias, depending on whether the true range is known at a prior time or not. Ranging accuracies of <1 m are possible with this method.

2. Introduction

The range to a spacecraft in geocentric polar coordinates, as determined by a range code, is given by (SPS 37-29, Vol. IV, pp. 229–239)

$$R = \int_{r_0}^{r_1} \frac{m(r)\, n_g(r)\, dr}{(r^2 n^2(r) - b^2)^{1/2}} \tag{1}$$

where $n(r)$ and $n_g(r)$ are the phase and group indices of refraction of the signal; r_0 and r_1 are the station and spacecraft radius vectors; and b is the impact parameter of the ray path.

We shall assume that the phase index of refraction has the form

$$n^2(r) = 1 + \epsilon f(r) \equiv 1 + \epsilon (f_1(r) - f_2(r)) \qquad (2)$$

where $|f(r)| \leq 1$ for $r_0 \leq r_1 < \infty$ and ϵ is a positive parameter. We shall consider $f_1(r)$ to be frequency independent, corresponding to the troposphere, and $f_2(r)$ to be frequency dependent, corresponding to the ionosphere and interplanetary medium. We shall consider frequencies at L-band (900 Mhz) or higher, in which case $\max f_1(r) > \max f_2(r)$ and $\epsilon = 2\alpha$ where α is the ground level refractivity (SPS 37-43, Vol. IV, pp. 314–321). If we neglect absorption and external magnetic fields the group index is given by (SPS 37-43, Vol. IV, pp. 314–321)

$$n_g(r) = \frac{1 + \epsilon f_1(r)}{n(r)} \qquad (3)$$

The range to a spacecraft in geocentric polar coordinates as determined by doppler is given by (SPS 37-29, Vol. IV, pp. 229–239)

$$R_d = \int_{r_0}^{r_1} \frac{rn^2(r)\,dr}{(r^2n^2(r) - b^2)^{1/2}} \qquad (4)$$

and the range rate determined by doppler is simply

$$\dot{R}_d = \frac{dR_d}{dt} \qquad (5)$$

A perturbation solution for the impact parameter is given by (SPS 37-43, Vol. IV, pp. 314–321)

$$b = b_0 + \epsilon b_1 + \epsilon^2 b_2 + 0(\epsilon^3) \qquad (6)$$

where

$$b_o = \frac{r_0 r_1 \sin\theta}{(r_0^2 + r_1^2 - 2r_0 r_1 \cos\theta)^{1/2}}$$

$$b_1 = \frac{1}{2}\,\frac{(r_1^2 - b_0^2)^{1/2}(r_0^2 - b_0^2)^{1/2}}{(r_1^2 - b_0^2)^{1/2} - (r_0^2 - b_0^2)^{1/2}}\,b_0 \int_{r_0}^{r_1} \frac{rf(r)\,dr}{(r^2 - b_0^2)^{3/2}}$$

$$b_2 = -\frac{(r_1^2 - b_0^2)^{1/2}(r_0^2 - b_0^2)^{1/2}}{(r_1^2 - b_0^2)^{1/2} - (r_0^2 - b_0^2)^{1/2}}\left[\frac{1}{2}\,b_0 b_1^2\left(\frac{1}{(r_0^2 - b_0^2)^{3/2}} - \frac{1}{(r_1^2 - b_0^2)^{3/2}}\right)\right.$$
$$\left. - \frac{1}{2}\,b_1 \int_{r_0}^{r_1} \frac{rf(r)\,dr}{(r^2 - b_0^2)^{3/2}} + \frac{3}{8}\,b_0 \int_{r_0}^{r_1} \frac{r^3 f^2(r)\,dr}{(r^2 - b_0^2)^{5/2}} - \frac{3}{2}\,b_0^2 b_1 \int_{r_0}^{r_1} \frac{rf(r)\,dr}{(r^2 - b_0^2)^{5/2}}\right] \qquad (7)$$

where θ is the angular separation between r_0 and r_1.

3. Elimination of Charged Electron Effects to First Order

In ranging a spacecraft we shall assume that at some time t_0 the range to the spacecraft is obtained by means of a range code; then the doppler shift is obtained and counted incrementally (in 1-sec intervals for example) until some time t_1, at which time the range is obtained by means of the range code. This process is then repeated from time t_1 to time t_2, from time t_2 to time t_3, etc.

Since we are incrementally counting doppler, this is equivalent to integrating Eq. (5) from time t_i to time t_{i+1}. Hence, the range change due to doppler count from time t_i to time t_{i+1}, from Eqs. (4) and (5), is

$$\Delta_i R_d \equiv \int_{t_i}^{t_{i+1}} \dot{R}_d dt = R_d(t_{i+1}) - R_d(t_i) \qquad (8)$$

The range change due to range code ranging from time t_i to t_{i+1} is, from Eq. (1),

$$\Delta_i R \equiv R(t_{i+1}) - R(t_i) \qquad (9)$$

A perturbation solution for $R(t_i)$ is (SPS 37-43, Vol. IV, pp. 314–321)

$$R(t_i) = R_s(t_i) + \frac{1}{2}\epsilon \int_{r_0}^{r_1(t_i)} \frac{r(f_1(r) + f_2(r))\,dr}{(r^2 - b_0^2(t_i))^{1/2}}$$

$$-\frac{1}{8}\epsilon^2\left[\int_{r_0}^{r_1(t_i)} \frac{r^3(f_1(r) - f_2(r))(f_1(r) + 3f_2(r))\,dr}{(r^2 - b_0^2(t_i))^{3/2}} - 2b_0(t_i)\,b_1(t_i)\int_{r_0}^{r_1(t_i)} \frac{r(f_1(r) + 3f_2(r))\,dr}{(r^2 - b_0^2(t_i))^{3/2}}\right] + 0(\epsilon^3) \qquad (10)$$

where $R_s(t_i)$ is the straight line or true range, and b_0 and b_1 are given by Eq. (7). We can obtain the same type of solution for $R_d(t_i)$ by substituting Eqs. (2) and (6) into Eq. (4) and expanding in powers of ϵ. This yields

$$R_d(t_i) = R_s(t_i) + \frac{1}{2}\epsilon \int_{r_0}^{r_1(t_i)} \frac{r(f_1(r) - f_2(r))\,dr}{(r^2 - b_0^2(t_i))^{1/2}}$$

$$+\frac{1}{8}\epsilon^2\left[2b_0(t_i)\,b_1(t_i)\int_{r_0}^{r_1(t_i)} \frac{r(f_1(r) - f_2(r))\,dr}{(r^2 - b_0^2(t_i))^{3/2}} - \int_{r_0}^{r_1(t_i)} \frac{r^3(f_1(r) - f_2(r))^2\,dr}{(r^2 - b_0^2(t_i))^{3/2}}\right] + 0(\epsilon^3) \qquad (11)$$

after utilizing Eq. (7).

If we add Eqs. (8) and (9) and divide by 2, we have, from Eqs. (10) and (11),

$$\Delta_i R_t \equiv \frac{\Delta_i R + \Delta_i R_d}{2} = R_s(t_{i+1}) - R_s(t_i) + \frac{1}{2}\epsilon \int_{r_0}^{r_1(t_{i+1})} \frac{rf_1(r)\,dr}{(r^2 - b_0^2(t_{i+1}))^{1/2}}$$

$$-\frac{1}{2}\epsilon \int_{r_0}^{r_1(t_i)} \frac{rf_1(r)\,dr}{(r^2 - b_0^2(t_i))^{1/2}} + \epsilon^2\left[\frac{1}{4}b_0(t_{i+1})\,b_1(t_{i+1})\int_{r_0}^{r_1(t_{i+1})} \frac{r(f_1(r) + f_2(r))\,dr}{(r^2 - b_0^2(t_{i+1}))^{3/2}} - \frac{1}{4}b_0(t_i)\,b_1(t_i)\right.$$

$$\left.\times \int_{r_0}^{r_1(t_i)} \frac{r(f_1(r) + f_2(r))\,dr}{(r^2 - b_0^2(t_i))^{3/2}} - \frac{1}{8}\int_{r_0}^{r_1(t_{i+1})} \frac{r^3(f_1^2(r) - f_2^2(r))\,dr}{(r^2 - b_0^2(t_{i+1}))^{3/2}} + \frac{1}{8}\int_{r_0}^{r_1(t_i)} \frac{r^3(f_1^2(r) - f_2^2(r))\,dr}{(r^2 - b_0^2(t_i))^{3/2}}\right] + 0(\epsilon^3)$$

$$(12)$$

The quantity $\Delta_i R_t$, as seen from Eq. (12), represents the range change due to spacecraft motion plus, to first order, the change due to the troposphere alone. The f_2-terms have disappeared from the ϵ-term which is our main achievement.

The coefficient of ϵ^2 in Eq. (12) consists of the difference of two terms of the form

$$\frac{1}{4}b_0 b_1 \int_{r_0}^{r_1} \frac{r(f_1(r) + f_2(r))\,dr}{(r^2 - b_0^2)^{3/2}} - \frac{1}{8}\int_{r_0}^{r_1} \frac{r^3(f_1^2(r) - f_2^2(r))\,dr}{(r^2 - b_0^2)^{3/2}} \qquad (13)$$

We can show that the above quantity multiplied by ϵ^2 is always less than 8 m for ranging in the solar system, excluding signals that pass close to the Sun. The upper limit of 8 m is obtained when the signal path grazes the horizon and the spacecraft is just outside of the troposphere. As the spacecraft moves farther away, the effect of the ionosphere and interplanetary medium reduces Eq. (13) in magnitude. Hence, we may neglect the coefficient of ϵ^2 in Eq. (12).

Since $f_1(r) = 0$ outside of the troposphere, we can write integrals of the form

$$T \equiv \int_{r_0}^{r_1} \frac{r f_1(r)\,dr}{(r^2 - b_0^2)^{1/2}} = \int_{r_0}^{r_2} \frac{r f_1(r)\,dr}{(r^2 - b_0^2)^{1/2}} \qquad (14)$$

where r_2 is the radius vector to the upper limit of the troposphere. Hence we can rewrite Eq. (12) (neglecting terms of $0(\epsilon^2)$ and higher)

$$\Delta_i R_t = R_s(t_{i+1}) - R_s(t_i) + \frac{1}{2}\epsilon T(t_{i+1}) - \frac{1}{2}\epsilon T(t_i) \qquad (15)$$

Since it can be shown (SPS 37-43, Vol. IV, pp. 314–321) that the coefficient of ϵ^2 in Eq. (10) is of the same magnitude as the coefficient of ϵ^2 in Eq. (12), we have (neglecting terms of $0(\epsilon^2)$ and higher)

$$R(t_i) = R_s(t_i) + \frac{1}{2}\epsilon T(t_i) + \frac{1}{2}\epsilon I(t_i) \qquad (16)$$

where

$$I = \int_{r_0}^{r_1} \frac{r f_2(r)\,dr}{(r^2 - b_0^2)^{1/2}} \qquad (17)$$

If the true range to the spacecraft is known at some range point at time t_0, then the range $\mathcal{R}$ at later times is obtained by means of Eq. (15), i.e.

$$\mathcal{R}(t_n) = R_s(t_0) + \sum_{i=0}^{n-1} \Delta_i R_t$$

$$= R_s(t_n) + \frac{1}{2}\epsilon T(t_n) - \frac{1}{2}\epsilon T(t_0) \qquad (18)$$

We see that the true range is obtained at t_n with tropospheric corrections at t_0 and t_n. If the true range is not known, then $\mathcal{R}(t_n)$ can be obtained with a constant bias. Upon combining Eqs. (15) and (17)

$$\mathcal{R}(t_n) = R(t_0) + \sum_{i=0}^{n-1} \Delta_i R_t$$

$$= R_s(t_n) + \frac{1}{2}\epsilon T(t_n) + \frac{1}{2}\epsilon I(t_0) \qquad (19)$$

which yields the true range with a tropospheric correction at t_n and a constant bias of ½ $\epsilon I(t_0)$. This bias can be minimized by choosing t_0 during the night and or when the spacecraft is at its highest elevation angle. As long as the spacecraft is outside of the influence of the troposphere, the corrections for the troposphere are not influenced by errors in the range to the spacecraft, which follows from Eq. (14).

4. Conclusions

The basic advantage of the ranging method presented in the article is that high precision ranging data free of charged electron effects can be obtained. The type of data required is precisely the type of ranging data that is obtained by tracking devices currently in use in the deep space net. The technique, which is very simple, could easily be automated at the tracking stations. Furthermore, since tracking is rarely done at low elevation angles, the error (one way) of 8 m can easily be reduced by a factor of 10; this is equivalent to tracking accuracies on the order of 10 nanosec.

References

1. Doelz, M. L., Heald, E. T., and Martin, D. L., "Binary Transmission Techniques For Linear Systems," *Proc. IRE*, Vol. 45, pp. 656–661, May 1957.

2. deHass, T., "Statistics of The Frequency-Differential Phase Relationship of CW Signals Transmitted Over a 4,000 Mile Ionospheric Path," *Proc. 10th National Comm. Symp.*, pp. 398-406, Oct. 1964.

Reference (contd)

3. Kathryn, H. F., *Radio Teletype And Data System Final Report—Phase I*, Report 938-230-14, Contract DA 36-039 SC-85192, General Atronics Corp., West Conshohocken, Pa., Sept. 1961.

4. Proakis, J. G., Drouilhet, P. R., Jr., and Price, R., "Performance of Coherent Detection Systems Using Decision-Directed Channel Measurement," *IEEE Trans. Comm. Systems*, Vol. CS-12, pp. 54–63, March 1964.

5. Bussgang, J. J., and Leiter, M., "Phase-Shift Keying With a Transmitted Reference," *IEEE Trans. Comm. Technology*, Vol. Comm-14, No. 1, pp. 14–22, Feb. 1966.

6. Van Trees, H., *Optimum Power Division in Coherent Communication Systems*, Technical Report 301, Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Mass., Feb. 1963.

7. Stiffler, J. J., "On The Allocation of Power in Synchronous Binary PSK Communication Systems," *Proc. 1964 National Telemetering Conference*, June 2–4, 1964.

8. Lindsey, W. C., "Phase-Shift Keyed Signal Detection With Noisy Reference Signals," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. AES-2, No. 4, pp. 393–401, July 1966.

9. Layland, J. W., *Signal Design For Communication With Timing Uncertainties*, Ph.D Thesis, Carnegie Institute of Technology, Pittsburgh, Pa., June 1965.

10. Costas, J. P., "Synchronous Communications," *Proc. IRE*, Vol. 44, pp. 1713–1718, Dec. 1956.

11. Stratonovich, R. L., *Topics in The Theory of Random Noise*, Vol. I, Gordon and Breach, Science Publishers, Inc., 1963.

12. Lindsey, W. C., "On The Design And Performance of Frequency Multiplexed, Phase Modulated, Block Coded Communication Systems," to be published in *IEEE Trans. Communication Systems.*

13. Viterbi, A. J., "On Coded Phase-Coherent Communications," *IEEE Trans. Space Electronics and Telemetry*, Vol. SET-7, pp. 3–14, March 1961.

14. Lindsey, W. C., "Optimal Design of One-Way and Two-Way Coherent Communication Links," *IEEE Trans. Communication Technology*, Vol. Com-14, No. 4, pp. 418–431, Aug. 1966.

15. Charles, F. J., and Lindsey, W. C., "Some Analytical and Experimental Phase-Locked Loop Results For Low Signal-to-Noise Ratios," Proc IEEE, Vol. 54, No. 9, pp. 1152–1166, Sept. 1966. Also see "A Model Distribution For the Phase-Error in Second Order Phase Locked Loops," by the same authors. *IEEE Trans. Communication Technology*, Vol. Com-14, No. 5, pp. 662–664, Oct. 1966.

Reference (contd)

16. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers,* Technical Report No. 32-819, Jet Propulsion Laboratory, Pasadena, California, Feb. 15, 1965.

17. Lindsey, W. C., "An Approach to the Determination of Modulation Indices and Designing Two-Channel Coherent Communication Systems," *IEEE Trans. Communication Technology,* April 1967.

18. Lindsey, W. C., "Phase-Shift-Keyed Signal Detection With Noisy Reference Signals," *IEEE Trans. Aerospace and Electronic Systems,* Vol. AES-2, No. 4, pp. 393–401, July 1966.

19. Marsh, R. W., *Table of Irreducible Polynomials Over GF(2) Through Degree 19,* distributed by Office of Technical Services, Commerce Department, Washington, D. C., Oct. 24, 1957.

20. Golomb, S. W., *Shift Register Sequences,* Holden-Day, Inc., San Francisco, California, 1967.

21. Burnside, W., *Theory of Groups of Finite Order,* 2nd Edition, Cambridge University Press, 1911 (Reprinted by Dover Publications, Inc., 1955).

22. Polya, G., "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und Chemische Verbindungen," *Acta Math.* Vol. 68, pp. 145–253, 1937.

23. Golomb, S., *A Mathematical Theory of Discrete Classification,* Proc. Fourth London Symposium on Information Theory, Butterworths, London 1961.

24. Hall, M., Jr., *The Theory of Groups,* MacMillan, New York, 1961.

25. Berlekamp, E. R., "Practical BCH Decoders," *IEEE Trans. Information Theory,* Vol. IT-13, to appear, 1967.

26. Mann, H. B., "On the Number of Information Symbols in Bose-Chaudhuri Codes," *Information and Control,* Vol. 5, pp. 153–162, 1962.

27. Gorenstein, D., Peterson, W. W., and Zierler, N., "Two Error-Correcting Bose-Chaudhuri Codes are Quasi-Perfect," *Information and Control,* Vol. 3, pp. 291–294, 1960.

28. Davenport, W B., Jr., "Signal-to-Noise Ratios in Band-Pass Limiters," *J. Appl. Phys.,* Vol. 29, No. 6, pp. 720–727, June 1955.

29. Eisenberger, I., and Posner, E. C., "Systematic Statistics used for Data Compression in Space Telemetry," *J. Amer. Statist. Assoc.,* Vol. 60, pp. 97–133, March 1965, and Technical Report No. 32-510, Jet Propulsion Laboratory, Pasadena, California, Oct. 1, 1963.

30. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient, I,* Technical Report No. 32-718, Jet Propulsion Laboratory, Pasadena, California, June 1, 1965.

31. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient, II,* Technical Report No. 32-755, Jet Propulsion Laboratory, Pasadena, California, Sept. 15, 1965.

Reference (contd)

32. Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946.

33. Trench, W. F., "An Algorithm For The Inversion of Finite Toeplitz Matrices," *J. Soc. Indust. and Appl. Math.*, Vol. 12, pp. 515–522, 1964.